

Perils of inadequacies in safety regulation

Last week's nuclear accident in Japan is a major blow to the nuclear power industry. But it also reflects broader problems in that country's management of technology.

Proponents of nuclear power must be shaking their heads in disbelief at last week's accident at a nuclear-fuel processing plant in Tokai, Japan. It seems that untrained and unprotected workers at the plant were blithely mixing excessive quantities of concentrated uranium solution in steel buckets with spoons in order to cut corners and speed up production. This led to a container of the solution going critical, exposed the workers to potentially lethal levels of neutron radiation, necessitated the evacuation of neighbouring houses and terrified the 310,000 citizens of Tokai as they were advised to take refuge in their homes and batten down the windows.

The nuclear accident at Tokai is just the worst of many in recent years. The responsibility lies squarely on the shoulders of the government and, more specifically, on those of the Science and Technology Agency, which is proving itself incapable of adequately regulating the safety of nuclear power. But the problem of the effectiveness of safety regulation in Japan is not confined to nuclear power — or the Science and Technology Agency. Similar deficiencies can, for example, be seen in Japan's regulation of the pharmaceutical industry, where drugs that are of questionable efficacy or downright dangerous — such as non-heat-treated blood products — are allowed onto the market.

The Japanese government seems unable to set up competent regulatory bodies with sufficient staff and expertise. The Science and Technology Agency's Nuclear Safety Commission is a group of part-time academic experts who rubber-stamp documents produced by a

small team of officials, who are far too few in number, and lack the expertise needed to regulate the safety of such a huge and potentially dangerous industry. Similarly, the country has no equivalent of the US Food and Drug Administration, even though its pharmaceutical market is of comparable size to that of the United States.

Will the situation improve significantly after this accident? Based on the record to date, probably not. When the Science and Technology Agency merges with the ministry of education in 2001, responsibility for most aspects of nuclear safety will probably pass to the Ministry of International Trade and Industry and/or a strengthened prime minister's office. But some responsibilities may remain with the merged ministry and agency. Unless the government commits adequate funding, manpower, expertise and accountability to a new regulatory body, the problems of the past will continue. However, there is little public pressure to do this in a society where respect for authority, despite its failings, remains high.

This is bad news for the world's nuclear industry. Despite calls in some circles for greater use of nuclear power to curb carbon dioxide emissions, the use of nuclear power is expected to decline in most of the developed world over the next 20 years. Only in Asia is significant expansion expected, in China, Japan, Korea, and to a lesser extent in Southeast and South Asia. But resistance to the siting of nuclear plants in Japan has been growing year by year, and the Tokai accident will no doubt pump opposition to new heights for many years to come. ■

Moment of truth for the test ban treaty

Republican senators will bring credit on their office if they consider the Comprehensive Test Ban Treaty on its technical merits, and vote for its immediate ratification.

For more than 40 years, nuclear weapons scientists and arms control advocates have sought an international agreement that would end the development and testing of nuclear weapons. The culmination of that effort, the Comprehensive Test Ban Treaty, was signed by 152 nations in 1996. The treaty is of particular benefit to the United States, which did more than any other nation to formulate its language and which holds a lead in both nuclear weapons design and the computer-simulation technology that will be used to study weapons in the absence of testing.

The US Senate, which must ratify the treaty with a two-thirds majority if it is to take effect, is debating its contents this week and may vote next Tuesday on its ratification (see page 519). During hearings before the vote, the scientific and military leadership of the United States will testify in favour of the treaty. Treaty critics claim that the US nuclear weapons laboratories' stockpile stewardship programme cannot assure the safety and reliability of the nuclear weapons stockpile in the absence of testing. But the laboratory directors, as well as independent technical advisers to the US government, believe that stockpile stewardship will work. Thirty-two US Nobel laureates in physics will this week release a letter expressing the same view.

The critics also contend that verification of compliance with the treaty will be impossible, as seismologists will be unable to identify small underground nuclear tests. The Central Intelligence Agency has just released an assessment stating that it will be unable to detect small, sub-critical explosions under the treaty regime. But sub-critical tests are permitted by the treaty. The American Geophysical Union and the Seismological Society of America will restate this week that the monitoring system to be established under the treaty is sufficient to meet its verification goals.

If the treaty is considered by the Senate on its merits, it should obtain the 67 votes needed for ratification by the United States — rapidly leading to a similar outcome in most, if not all, of the 44 nations that must ratify it before it takes effect. Alas, the timing and circumstances of the vote put pressure on Republican senators to oppose the treaty on partisan grounds. Some Democrats, much to their discredit, are banking on this so that they can portray the Republicans as proponents of nuclear war during next year's elections. Republican senators such as Richard Lugar (Indiana) and Pete Domenici (New Mexico) must confound them by rising above partisan considerations on Tuesday and voting to ratify the treaty. ■





Mars mission
'Full conversion to metric system might have saved Orbiter'
p517



Ig Nobels
Heredity was the dominant theme at this year's awards
p518



Rice genome
US sequencers to receive funding for plant genomes
p520



Australia
CSIRO braces for shift to demand-led research strategy
p521

Terms of access to cloned mice comes under researchers' fire

San Diego

A powerful 'gene-trapping' technology, first described in a paper published in *Nature* last year, has become the focus of a battle for access to proprietary molecular tools, with academic researchers pitted against the company that owns the technology.

The dispute, which sets researchers from universities and the Howard Hughes Medical Institute (HHMI) against the privately owned Lexicon Genetics of Texas, has highlighted a growing source of tension between the academic world and corporate interests. Academics want reasonable access to discoveries published in the scientific literature, while companies such as Lexicon seek to profit from their proprietary technology.

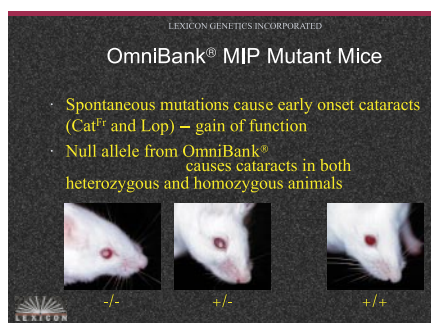
Some academics say the case illustrates the need for stricter rules to make researchers — whether from academia or industry — who publish in the open literature provide reagents and molecular tools to non-commercial investigators without restrictive provisions in material-transfer agreements.

In April 1998, six scientists from Lexicon published an article (*Nature* 392, 608–611; 1998) describing a high-throughput mutagenesis method in which 'gene trapping' provides an automated way of identifying sequence tags from mutated genes.

The Lexicon scientists reported that they were creating a library of mouse embryonic stem cells, called Omnibank. They noted that sequence-tagged mutations in 2,000 genes had initially been placed in Omnibank. The library now includes around 70,000 mouse genetic clones. The paper concluded: "Lexicon will distribute the [embryonic stem] cell clones described here to requesting investigators for non-commercial research."

But when Lawrence S. B. Goldstein, an HHMI geneticist at the University of California at San Diego, tried to obtain a particular mutated embryonic stem cell line last summer, he ran into what he termed "onerous" terms that became a "roadblock" to receiving the cell line. Lexicon sought at least \$15,000 for the stem cells, Goldstein says, as well as 'reach-through rights' to any commercial product from his use of the cells.

Pointing out that he believed his planned



Mouse trap? Some scientists claim Lexicon's embryonic stem-cell library has 'excessive' costs.

experiments "had no commercial value", Goldstein describes Lexicon's open publication of the technique and subsequent restrictions as a type of "shell game" in which the company used the journal for "an advertisement". Goldstein says he has no objections to a company placing tough restrictions on its unpublished materials. But he adds: "If you aren't going to send materials out, don't publish and take the credit."

Michael Green, a HHMI molecular biologist at the University of Massachusetts

Medical School at Worcester, said he had a similar experience when asking Lexicon for a retroviral vector referenced in the same paper. After his inquiries to the company had gone unanswered, he says, he obtained a response from Lexicon only by enlisting the assistance of a *Nature* editor.

But Lexicon then set restrictions on the transfer of the vector that neither HHMI nor his university would agree to, says Green. After months of delays, he finally received an ill-defined description of the sequence for the vector. But because of the insufficient quality of the sequence description and the delay, Green reluctantly used another vector obtained from an academic researcher.

Arthur Sands, the chief executive and a founder of Lexicon, denies that his company has violated *Nature's* policy on sharing materials. A physician with a PhD in molecular biology, and a co-author of the paper, he says the company "operates in good faith" and has "very liberal agreements" on sharing resources. "We have distributed clones from the publication very broadly," he says.

Sands attributes Goldstein's difficulties to a "misunderstanding" and "confusion" ▶

UK marine centres face job losses

London

Forty-nine research and technical jobs are to be lost at Britain's Centre for Coastal and Marine Sciences (CCMS), because of what its core funding agency, the Natural Environment Research Council (NERC), describes as "severe financial difficulties".

Twenty-seven redundancies will be compulsory. Most will come from the Plymouth Marine Laboratory, and four from the Proudman Oceanographic Laboratory.

The redundancies have been blamed on declining support from government and industry. Over the past three to four years, CCMS has seen a 30 per cent decline in government contracts and a deterioration in support for long-term research.

The Institute of Professionals, Managers and Specialists — the union that represents many research council scientists — has expressed concern over the job losses.

But it welcomed a £7 million (US\$11.6 million) rescue plan from NERC. This will cover the centre's deficits from the past two years, and the estimated cost of the redundancies.

NERC has approved a package that cancels an extension to the Plymouth Marine Laboratory and privatizes some of its work. After a review of CCMS science programmes, it has been decided that redundancies will occur in the research areas of lipid membranes, larval fish biology, numerical modelling and southern ocean dynamics.

Natasha Loder

regarding material-transfer agreements, although when contacted by *Nature* he said he was unaware of Green's claimed difficulties. Sands says that he subsequently contacted Goldstein to try to "clear up any misunderstanding", but at the time of going to press Goldstein said he had not received the material.

Sands says that, since the publication of the paper, his company has transferred materials to 41 non-commercial investigators at 27 institutions. He declined to provide a list of the investigators, but named half a dozen researchers at four universities. Interviews with some of these showed a substantial fee was paid for unpublished material, while one institution was asked to cede reach-through rights to Lexicon.

These concerns prompted HHMI officials to circulate a memorandum to its 300 researchers. HHMI spokesman Robert Potter issued a statement: "Several HHMI investigators have expressed interest in having access to Omnibank and related resources owned by Lexicon for their research. So far, we have been unable to come to terms on an arrangement for such access."



Goldstein: faced with "onerous" terms.

To researchers, the conflict shows that enforcement is needed to make material from the published domain more readily available. Such problems remain, says Green, because "no one wants to take responsibility for the enforcement".

Green argues that the National Institutes of Health, which last spring issued suggested guidelines for making materials available from its own funded research (see *Nature* 399, 291; 1999), should take a greater responsibility for the situation. NIH director Harold Varmus, a champion of ready access to published research materials, was unavailable for comment.

Several scientists said major journals that have policies on access to materials from published research should prohibit authors from publishing if they break the rules on material transfers. **Rex Dalton**

Nature's policy is that materials should be made freely available. This is made explicitly clear to all authors as a condition of publication. Where the conditions are subsequently broken, *Nature* reserves the right, as one possible sanction, to refuse to consider further papers from the authors or even, where necessary, the institution or company concerned. We are contacting the authors at Lexicon and will inform readers of the outcome. **The editor**

Company to use advertising to cover Pubmed Central costs

Baltimore

A private company that provides researchers with information about funding opportunities and other activities announced last week that it will provide 'front end' services on Pubmed Central, the free repository for research results which the National Institutes of Health (NIH) plans to launch in January.

These services will enable societies and individuals to publish peer-reviewed research and are to be provided by Community of Science (COS), based in Baltimore, which keeps profiles of 500,000 researchers globally.

Huntington Williams, COS's president said his company was trying to formulate an economic model for publishing on Pubmed Central. The costs of processing and reviewing manuscripts would be covered by online advertising and direct marketing aimed at the scientists who select the reviewers and the reviewers themselves.

COS is the first organization to propose a business model that would allow electronic journals which publish on Pubmed Central — and therefore have no subscription revenue — to cover the costs of arranging for the review of scientific papers, and editing the text and illustrations into a standard format ready for publication.

These costs would be covered by selling web advertising and web marketing targeted at the reviewers themselves and at the boards of researchers that select them. The existing COS database would enable such advertising and marketing to be tightly targeted at these scientists' interests and personal habits.

Williams refused to speculate over how much money could be raised in this way, but

I've just agreed to review someone's research, bought a new centrifuge, and booked a week's holiday in Beautiful Baltimore...



the newsletter *Science and Government Report* quotes COS officials as saying that, while unspecified journals spend \$4000 to process each paper, the new system might do it for \$250. The model assumes that this amount, plus profit, could be generated by advertising and marketing aimed at referees.

Williams said the model would enable COS subscribers, as well as societies who wanted to publish their journals on Pubmed Central, to publish results in the new repository. COS and the societies would revenues from the advertising and marketing, he said.

David Lipman, director of the National Center for Biotechnology Information at the NIH and one of the architects of Pubmed Central, said that COS's plans were one of "a wide variety of business models" which the repository would accommodate. He adds that plans for PubMed Central's launch are focusing mainly on the part incorporating peer-reviewed content. **Colin Macilwain**

Optical society vote sees off merger

San Diego

Members of the Optical Society of America (OSA) voted last week not to merge with the International Society for Optical Engineering (SPIE), ending months of contentious campaigning at the two societies.

The result was announced last week at the society's annual meeting in Santa Clara, California. SPIE had voted by mail earlier in the summer, with the result to be announced on Tuesday of this week. But the response from the OSA membership kills the merger, which was first mooted in early 1998.

OSA's leadership had sought the merger to create a more powerful organization. But criticism of the merger proposal grew as vocal dissidents expressed concerns that the

research-orientated OSA might be threatened by SPIE and its focus on applied research (see *Nature* 398, 547; 1999).

The OSA voted against the merger by 51 per cent (2,551 votes) to 49 per cent (2,420 votes). A two-thirds majority of voting members was needed to approve the merger.

"The members have spoken," said Anthony E. Siegman, OSA's president and a leading proponent of the merger. "In their view, a merger of OSA and SPIE is not in the best interests of the society at this time."

Daniel V. F. James, a theoretical physicist at Los Alamos National Laboratory in New Mexico and a critic of the merger proposal, said of the vote: "We are happy; it was the correct decision." **Rex Dalton**

NASA reworks its sums after Mars fiasco

Washington

Confusion between imperial and metric units was to blame for the loss of the US space agency NASA's Mars Climate Orbiter, according to project managers at the Jet Propulsion Laboratory.

The spacecraft was lost on 23 September, when a targeting error sent it into, rather than safely above, the Martian atmosphere as it was about to enter orbit (see *Nature* **401**, 415; 1999). The spacecraft was either destroyed or skipped off the atmosphere into an unknown orbit around the Sun.

A preliminary analysis has revealed that the builder of the spacecraft, Lockheed Martin Astronautics of Denver, had been supplying information about routine propulsion manoeuvres in imperial units. But NASA space navigators assumed throughout the nine-month voyage from Earth to Mars that the acceleration data were in metric units, resulting in a large targeting error when the orbiter reached its destination.

The spacecraft's operators performed a major manoeuvre to correct its trajectory a week before it reached Mars, and thought they were on course at that point. But subsequent, smaller propulsion manoeuvres were required every 15 hours or so to counteract the energy built up in onboard 'reaction wheels' used to orientate the spacecraft.

Errors in these routine smaller manoeuvres are believed to have resulted in the large deviation when the spacecraft reached Mars. Three separate review teams are currently examining the mishap in detail, with reports due in mid-November.

The first priority for NASA is to ensure that its Mars Polar Lander, also built by Lock-



Getting it right? NASA hopes not to confuse metric and imperial for its Mars Polar Lander

heed Martin and due to touch down on the planet's surface on 3 December, does not have the same problem.

A trajectory correction planned for the lander early this month has been postponed until project engineers have a better grasp of the situation, says Noel Hinners, a vice-president at Lockheed Martin, which monitors the spacecraft from its operations centre near Denver. "We don't want to do anything until we understand this one," he says.

Unlike the Mars Climate Orbiter, the lander does not use reaction wheels to control its altitude during the interplanetary cruise, according to Hinners, so the same problem is unlikely to occur. But having made one costly mistake, Lockheed Martin is carefully

scrutinizing the two missions it currently operates for NASA — the Mars Polar Lander and the Stardust comet sample return — to see if "there are any other things that could have slipped through".

The company has long experience of controlling NASA spacecraft, dating back nearly a decade to the Magellan Venus mission. But a recent external review of its Space and Strategic Missiles Sector, of which the astronautics division is a part, found systematic problems with quality control, with overemphasis on cost savings, and with management of subcontractors and parts suppliers.

The confusion between imperial and metric that doomed the spacecraft originated with a Lockheed Martin subcontractor, says Hinners. The supplier of the spacecraft's propulsion system provided accompanying data for its system in imperial units, and Lockheed Martin simply incorporated those data and passed them through the system.

"We should have caught [the mistake] then and there," he says. The propulsion system for Lockheed Martin's Mars Global Surveyor, which is now operating successfully in Martian orbit, was built by another subcontractor who used metric units.

The embarrassing mistake also underlines the lack of uniformity in the US space sector. While virtually all scientists use metric units, many US engineers, both inside and outside the space programme, use imperial units, converting them when necessary. If the United States had converted to the metric system when the subject was debated more than 20 years ago, muses Hinners, "it might have saved this mission". **Tony Reichhardt**

Virus treatment questioned after gene therapy death

San Francisco

Researchers at the University of Pennsylvania are investigating the first death in a gene therapy experiment, which was revealed last week. Their enquiries centre on the adenovirus vector used to deliver potentially therapeutic DNA to the liver.

Jesse Gelsinger, an 18-year-old, high-school graduate from Arizona, developed a fever and blood clots throughout his body within hours of treatment to correct partial ornithine transcarbamylase (OTC) deficiency, a rare metabolic disease that can cause a dangerous build-up of ammonia in the body. He died four days later.

US officials immediately began notifying the 100 or so gene therapy experimenters using adenovirus vectors, which are made

using a disarmed version of the virus that causes the common cold. Both the Food and Drug Administration (FDA) and National Institutes of Health (NIH), however, stopped short of recommending any policy changes or clinical holds.

According to the protocol, researchers at the University of Pennsylvania's Institute for Human Gene Therapy used an 'E1-deleted, E2A-temperature-sensitive' adenovirus vector to infect liver cells with the normal OTC gene, which codes for a urea-cycle enzyme that removes excess nitrogen from the body.

Gelsinger, the eighteenth and final patient in the Phase I experiment, was the second person to receive a dose of 3.8×10^{13} virus particles, believed to be the highest so far with an adenovirus. The virus was

delivered by a catheter inserted into the groin artery and advanced into the vessel that feeds the liver.

Researchers at the university are reviewing their notebooks and toxicology data. They are studying the vector, treating it with restriction enzymes and testing it in primates.

They have also conducted an autopsy and are examining tissues and liver cells, looking in particular for vector-related problems because of the speed of Gelsinger's reaction. "So far, we haven't seen anything that we'd do drastically different," said Nelson Wivel, deputy director of the University of Pennsylvania's Institute of Gene Therapy.

Inder Verma, professor of genetics at the Salk Institute in La Jolla, California, praises the Pennsylvania team for their openness, ▶

saying it will help to preserve public confidence and allow scientists to learn from the incident, "I'm sure it will introduce a note of caution in every experimentalist who does gene therapy, and that's a good thing," he says.

Verma describes vectors as the "Achilles heel" of gene therapy, and says that dose-escalation studies using adenoviruses should be re-examined. Most gene therapy involves retroviral vectors, but adenoviruses are popular for cancer and cystic fibrosis.

Verma says he thinks vectors like the one used in the OTC trial will soon be abandoned in favour of 'gutless' adenoviruses, retroviruses, AAV or lentivectors.

A major problem of adenoviruses is that even inactivated versions can stimulate an immune response. Sustained expression of the gene is therefore impossible, and the immune system may destroy infected cells — the very cells targeted for help. The severe inflammation associated with these vectors, especially in the liver, is particularly dangerous for OTC patients.

Members of the NIH's Recombinant DNA Advisory Committee (RAC), which at



Verma: new caution is "a good thing".

the time held regulatory authority over gene therapy experiments using federal funds, approved the protocol 11 to one, with four abstentions. They raised concerns about the risks of the treatment and its use — for the first time in gene therapy — in asymptomatic patients.

Individuals without the targeted gene cannot break down nitrogen, which can lead to a fatal build-up of ammonia soon after birth. The gene is X-linked — female carriers usually lead normal lives, but up to ten per cent of them could experience dangerous symptoms. Patients with partial enzyme production have done well under dietary and drug treatment.

The investigators had studied their vector in mice and primates. In the December 1995 review of the proposal for a trial in humans, RAC members discussed the potential for lethal liver inflammation

based on toxicity results in Rhesus monkeys and one animal's death after an extremely high dose of a first-generation vector.

The likelihood of efficacy and the importance for this and other liver conditions convinced them to approve the study, with the recommendation that the researchers use a less invasive route of administration through a peripheral vein.

Because of concerns about infection of reproductive cells, FDA regulators made the researchers go back to treating the liver directly. Under current rules, such an experiment might be reviewed by the RAC, but only requires approval from the FDA. An FDA spokesman said that the agency cannot discuss experiments or their results without the consent of the investigators concerned.

Verma and others point to the technology's success in thousands of patients up to now, and say that the death should not be seen as a setback for gene therapy as a whole. "We would obviously prefer that this tragedy had not happened," says Wivel. "But these things do occur in cutting-edge research."

Sally Lehrman

Ig prizes spawn a new generation of Nobels

Boston

Heredity was the theme of the 'Ninth First Annual' Ig Nobel Prize ceremony at Harvard University last week, where the King and Queen of Swedish Meatballs presided as usual over the festivities.

Beach-balls, paper aeroplanes and giant doughnuts sailed through the air of a Harvard University auditorium, while Lawyers For and Against Heredity carried placards and a group of female physics students called Babes in Boyland marched in procession. In keeping with the heredity motif, descendants of famous scientists, including Francis Crick's granddaughter, took the stage.

The world premier of *The Seedy Opera*, a tribute to scientist and entrepreneur Richard Seed (winner of the Ig Nobel Prize for Economics in 1998), who has vowed to clone himself and other humans, was presented in four acts. Four Harvard Nobel laureates — Sheldon Glashow, Robert Wilson, Dudley Herschbach and William Lipscomb — played the parts of cloned sheep in the opera.

This year, Ig Nobel prizes were awarded in ten categories for work that either "cannot or should not be reproduced". Recipients came from as far afield as Australia, Japan, Norway and the United States to collect the coveted Ig prize, which resembles a plaster frog.

The science-education prize was issued jointly to the Kansas and Colorado boards of education for mandating that "children should not believe in Darwin's theory of evo-



Cloning around: the world premier of *The Seedy Opera*, a tribute to one of last year's winners, Richard Seed, the Chicago-based former physicist who has vowed to clone himself and others.

lution, any more than they believe in Newton's theory of gravitation, Faraday's and Maxwell's theory of electromagnetism, or Pasteur's theory that germs cause disease."

Commenting on the award, University of Kansas biologist Douglas Ruden warned other states "not to let what happened in Kansas happen to you. As Dan Quayle said: 'A mind is a terrible thing to lose'."

Len Fisher of England and Australia captured one of two physics prizes, for calculating the optimal way to dunk a biscuit (see *Nature* 397, 469; 1999). The other award went to Jean-Marc Vanden-Broeck of the University of East Anglia for calculating how to make a teapot spout that does not drip.

Norway's Arvid Vatle earned the prize for medicine for an analysis of the containers chosen by patients for urine samples. The late

George and Charlotte Blonsky were posthumously awarded the managed health care prize for inventing a high-speed rotary chair for women in labour, to accelerate the birth process. "As chair of preventive medicine at Brigham and Women's Hospital, it's my job to prevent this kind of medicine," said JoAnn Manson of Harvard Medical School.

Takeshi Makino of Japan won the chemistry prize for developing an infidelity-detection spray for wives to apply to their husbands' underwear. The peace prize went to South Africa's Charl Fourie and Michelle Wong for inventing a car burglar alarm that comes equipped with a flame-thrower. "These achievements speak for themselves," claimed master of ceremonies Marc Abrahams, editor of the science humour magazine *Annals of Improbable Research*.

Steve Nadis

Bigger stipends attract more UK postgrads into physics

London

The number of British graduates signing up for doctorates in physics and engineering has increased for the first time in three years. The rise could be in response to last year's £1,000 (US\$1,600) increase in the value of the PhD stipend, bringing the minimum annual pay to £6,500.

The increase will be especially welcomed by the Engineering and Physical Sciences Research Council (EPSRC), which has made offers on 92 per cent of its research studentships. Last year it filled only 85 per cent, and 560 studentship (including masters) places went unfilled — more than some councils actually offer (see *Nature* **397**, 635; 1999).

The figures suggest that financial considerations have been responsible for the decline in interest in postgraduate research. Provisional figures on degree grades also indicate an increase in the qualifications of graduates taking on studentships in particle physics and astronomy (see table).

Providing staff to industry and academia is increasingly seen as a core function of the research councils. If confirmed across all

Increase in take-up of UK research studentships 1998 to 1999

Research council	1998		1999	
	Take-up	Firsts	Take-up	Firsts
EPSRC	85%	41%	92.3%	40.1%
PPARC	91.1%	48%	97.6%	57%

1998 figures are final; 1999 figures are provisional.

councils, the increase will be welcome news, as the councils are currently preparing to report to the government on how well they are managing their budget, to help determine the next round of funding.

John Taylor, director-general of the research councils, has already told the councils that his Office of Science and Technology's funding proposals will focus on 'world class science', 'people related' output, infrastructure, exploitation, and emerging priorities such as genomics and information technology.

The EPSRC figures remain equivocal. But take-up at the Particle Physics and Astronomy Research Council (PPARC) is now at 98 per cent, with nearly 60 per cent of students holding first-class degrees.

Natasha Loder

Test ban treaty faces a rough ride from Senate Republicans

Washington

US arms control advocates are concerned about whether they can secure anything close to the necessary two-thirds majority in next Tuesday's Senate vote on ratifying the Comprehensive Test Ban Treaty (CTBT).

Supporters of the treaty, including the American Physical Society, the Union of Concerned Scientists and several other scientific organizations, say that the hurried vote, only announced by the Senate leadership last week, makes it unlikely that many more Republican senators will join the two — Arlen Specter (Pennsylvania) and James Jeffords (Vermont) — who are supporting the treaty.

Thirty-two Nobel laureates in physics are expected to release a letter supporting the CTBT today (7 October). A defeat could set the treaty back several years, but some Democrat senators are keen for a vote, so they can attack Republicans who oppose a treaty supported by more than 80 per cent of voters.

Colin Macilwain

Japan joins efforts to patent cDNA clones

Tokyo

A Japanese genomics company, funded jointly by government and industry, has filed patent applications for more than 6,000 full-length human complementary DNA (cDNA) clones from various tissues, including the ovary, placenta and brain.

If it is successful, the Helix Research Institute plans to seek similar patents in the United States and Europe. But industry analysts say the Japanese Patent Office may be reluctant to award patents, pointing out that, although 700 of the clones are thought to correspond with genes that express membrane and secretory proteins, this alone would not satisfy the criteria of a patented invention. The functions of the remaining cDNA clones have not been specified.

The daily newspaper *Yomiuri Shinbun* reported last week that Helix — a joint venture between the Ministry of International Trade and Industry and ten private companies — applied for the patents in July. Helix hopes to produce, and possibly patent, a fur-

ther 20,000 full-length human cDNA clones over the next two years, in collaboration with other companies and research institutes, possibly through a consortium.

Helix's move reflects its concern over growing competition from US biotechnology ventures, such as Incyte, which are accelerating their efforts to clone full-length human cDNAs.

In a bid to preserve open access to this information, the US National Institutes of Health is creating a public repository for full-length human cDNAs to provide the research community with genetic sequences and clones (see *Nature* 401, 418; 1999).

Japan's cDNA project, which is led by Tokyo University, is aiming to sequence 30,000 cDNA clones — more than 20 per cent of expressed human genes — by 2001.

But Japan is falling behind the rest of the world in patenting genes, according to Hitoshi Watanabe, an executive director of Helix. A sense of urgency among researchers and industry has led to the government's decision to step up support for

biotechnology (see *Nature* 400, 389; 1999).

"It is important to secure patents [on human genes] for future development of new pharmaceutical products as well as biomedical research," says Watanabe.

Watanabe declines to elaborate on the filed patents, other than saying that the cDNAs were divided into several groups. But sources close to the company confirm that three patents were filed, one of which is for a set of 200 full-length cDNA clones, corresponding to genes that express membrane and secretory proteins.

The second patent is for a set of 850 full-length cDNA clones, 500 of which are also believed to express membrane and secretory proteins. The third is for the remaining 5,000 or so full-length cDNAs, although their functions are not known.

"We wouldn't have filed the patents if we thought there was no chance of them being approved," says Watanabe. But he admits he does not know how the applications will fare, as this is the first attempt to patent full-length human cDNAs in Japan. **Asako Saegusa**

US provides funding for sequencing rice genome

Washington

The US government has awarded \$12.3 million to The Institute for Genomic Research (TIGR) in Rockville, Maryland, and a consortium of universities to finance the country's component of an international effort to sequence the rice genome.

TIGR will receive \$7.2 million over three years, with another \$5.1 million going to Clemson University in South Carolina, Cold Spring Harbor Laboratory in New York, and Washington University in St Louis, Missouri.

Between them, the institutions plan to sequence 50 million of the 430 million DNA base pairs estimated to make up the rice genome. They will begin with the 24 million bases in chromosome 10, which has been allocated to the United States.

Although the rice genome was initially due to be completed by 2008, Japan, which is leading the project, has announced that it is planning to provide additional funding to complete its own sequencing efforts by 2004 (see *Nature* 401, 102; 1999).

Ed Kaleikau, director of plant research at the competitive grants office of the US Department of Agriculture (USDA), said that the US push would follow from the existing effort to sequence the genome of the model plant *Arabidopsis*, due to be completed by the end of next year.

"We're using a standard approach that



Take your pick: US efforts may focus on the indaco strain that is prevalent outside Japan.

builds on the strengths of the human and *Arabidopsis* [genome sequencing] projects," he says. USDA and the National Science Foundation (NSF) will each contribute \$6 million to the project, with another \$300,000 coming from the Department of Energy.

The international project will use a conventional approach to find the sequence of japonica rice, which is a commercially important strain.

Celera, the Rockville-based gene-sequencing company run by Craig Venter, has boasted that it could sequence the rice genome more rapidly than the publicly funded project, using its 'shotgun' approach (see *Nature* 398, 545; 1999). But a spokesman for Celera says that it is likely to concentrate on the genome of indaco rice, another important strain.

Indaco is the prevalent strain grown outside Japan and, according to Kaleikau, China has said that it will study it, rather than japonica rice, when it sequences chromosome 4 of the rice genome for the project.

Earlier this month, the NSF announced \$60 million of research grants under the second year of its plant genome project. These show that the agency plans to support genomics research in a wide range of commercially important plants, including multi-million dollar awards for research on potatoes and loblolly pine trees, as well as wheat, maize and *Arabidopsis*.

The NSF also said it would support a \$5.3 million *Arabidopsis* information resource, to be hosted jointly by the Carnegie Institution's Department of Plant Biology at Stanford University, California, and the National Center of Genome Resources at Santa Fe, New Mexico, which will disseminate data on the plant's genome. **Colin Macilwain**

CSIRO faces a 'demand-led' future

Australia's main research agency has responded to the prospects of continuing budgetary stringency by redefining its strategic approach.

Sydney

Australia's most powerful research organization, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), is shifting resources and staff away from fundamental research towards the 'strategic' and applied end of the research spectrum.

Job losses are expected in disciplines no longer considered a priority and in research groups not considered suitable for redeployment to other areas. The total staff of CSIRO is close to 7,000, but in four of its 23 divisions — telecommunications/industrial physics, forestry, building and food science — management has already said 89 jobs are to be cut.

CSIRO officials argue that "scope for some purely curiosity-driven research" will remain. But apart from radioastronomy, in which Australia continues to derive considerable international prestige, the days of broad support for basic research appear to be over.

The main goal of CSIRO will now shift to supporting "wealth creation and environmental sustainability for Australia". Research will be run like an investment portfolio, and the growth of research will be 'demand-led' by clients and research collaborators.

Government stringency

The organization, set up in 1926, has been unusual in maintaining a brief across most research disciplines (apart from medical research) and in blending basic and applied science. In 1991, its then sister body in New Zealand was split into more specialized units, but CSIRO has resisted such moves.

How long this can be sustained remains to be seen as its effort becomes increasingly determined by commercial liaisons, products and services that are favoured by government but protected by confidentiality agreements.

But CSIRO's executive committee has accepted that the government is unlikely to reverse the budget stringency it has imposed over the past four years. According to a plan for 2000–2003, released to staff by chief executive officer Malcolm McIntosh shortly before he went on sick leave last month for a cancer operation, its research strategy will be increasingly driven by commercial considerations.

CSIRO's research is currently performed in 23 specialist divisions. But recently a parallel network of 23 different sectors grouped in five 'alliances' has been established to coordinate multidisciplinary attacks on strategic problems, such as the increasing salinity of agricultural land.

The new plan presents the first compre-



Jobs for burning: it has proved difficult to get the forestry industry to invest in research.

hensive test of this reorganization. Each sector draws on research from several divisions, briefed by advisory committees, with strong input from industry, that contributed to the most intensive internal review of the organization's strategy for several years.

CSIRO's operations are financed from an annual government appropriation of A\$450 million (US\$295 million), augmented by income from contracts with private industry and other public sources, which increased its total budget to A\$700 million last year.

The government previously set CSIRO a target for external earnings of 30 per cent of its block grant. But the organization exceeded this, achieving 33 per cent last year and aiming for an average of 40 per cent next year across all divisions and sectors.

In defending the changes, Colin Adam, a metallurgical engineer acting as chief executive during McIntosh's absence, says that basic research should be focused in universities. In contrast to the government's green paper earlier this year, he also expresses doubt over the government's expectations that universities can significantly boost their earnings from industry (see *Nature* 400, 703; 1999).

"Universities should be heavily funded by taxpayers," says Adam. But he admits that tapping substantially into private resources will be "a big task", given that industry has been decreasing its own research and develop-

ment (R&D) following the government's controversial cut in a tax incentive.

Increasing competition ahead

In turn, university leaders, who did not play an institutional role in the CSIRO review, are expecting increasing competition between CSIRO and academic researchers for industry funding.

Peter Cullen, president of the Federation of Australian Scientific Technological Societies and professor of resource and environmental science at the University of Canberra, welcomes CSIRO's determination to set a strategy and be selective. But he foresees "a bitter environment" over price competition as "CSIRO and universities battle in the face of falling industry investments in R&D".

CSIRO's plan states that any area of research that fails to attract sufficient support from industry should be dropped. Applying this criterion, major cuts in the allocation of funds will reduce research in forest products, biodiversity, meat, dairy, aquaculture, telecommunications and industrial physics.

Criticism from staff is expected, and there has already been a protest at the forest products section of the forestry division, where 37 of the 83 staff in its Melbourne laboratory are to be made redundant. But any murmurings of discontent are likely to be subdued. Peter O'Donoghue, secretary of the union for CSIRO staff, claims staff feel reluctant to speak out against the plan.

Adam says it was "almost impossible" to get the forestry industry to invest in research and that this sector therefore had to be reduced. By contrast, allocations to exploration, mining and minerals, which Adam manages, have risen as they have attracted stronger industry support.

CSIRO's executive committee has agreed to give the biodiversity sector, which draws on work from eight divisions, a second chance to argue against what its researchers feel is a raw deal resulting from A\$4.9 million in cuts. Brian Walker, chief of the division of wildlife and ecology, says this shows the challenge facing those who are unable to persuade the executive that a lack of support from "customers and stakeholders" can be corrected.

Walker describes this as the classic "dilemma of the commons", arguing that the private sector seems unaware of the risks to the productivity of its land as biodiversity diminishes, and so is unwilling to pay for research to protect its sustainability. Others are waiting to see whether CSIRO's broader strategy can avoid a similar fate.

Peter Pockley

Research will be
run like an
investment portfolio

Senate proposes \$2bn increase in NIH funding

Washington Members of the US Senate appropriations committee have proposed a \$2 billion increase in funding for the National Institutes of Health (NIH) next year, which would increase the agency's budget to \$17.6 billion. The Senate was expected to pass an appropriations bill containing the mammoth increase earlier this week.

The House of Representatives is considering a bill that would give the NIH an increase of \$1.35 billion. But the means by which it has found the money — including the deferral of tax rebates to the working poor — has raised criticism that may prevent the bill from passing. The two proposals indicate that the NIH is poised to receive another substantial funding increase next year.

Rival amendments to the appropriations bills, which would have modified the existing ban on federally funded embryo research to permit or to explicitly prohibit the derivation of stem cells from embryos, were withdrawn after Congressional leaders made it clear that they did not want the controversial issue to further complicate the passage of the spending bills.

Researchers hit the bottle in drinking-water scare

London Researchers at a new department of cardiovascular medicine, a joint project between the University of Cambridge and the pharmaceutical company SmithKline Beecham, are being provided with bottled water following concerns over the contamination of drinking water.

The move was made after a routine inspection uncovered high levels of lead and iron in the unit's water supply. Researchers have been drinking bottled water since the beginning of August, although investigations have been hampered by officials' inability to replicate the findings. A spokesperson for the company said it hoped to complete its enquiry in the next couple of weeks.

Arima forced out in Japanese reshuffle

Tokyo Akito Arima, Japan's minister of education and director-general of the Science and Technology Agency (STA), has been replaced in a reshuffle of prime minister Keizo Obuchi's cabinet, which took place on 5 October. The post goes to Hirofumi Nakasone, the son of former prime minister Yasuhiro Nakasone who set up the Human Frontiers Science Program.

The reshuffle reflects Obuchi's re-election promises. Arima, a theoretical physicist and former dean of Tokyo University (see *Nature* 393, 300; 1998), had hoped to take an active role in the planned merger of the STA and the education ministry in 2001. The clean-up of last week's nuclear accident at a uranium-processing plant in Tokai Village, Ibaragi prefecture, became his last job as director of the STA.

US energy department loses science office head



Washington Martha Krebs (left), assistant secretary for science programmes at the US Department of Energy, will leave government in December to pursue other interests. For the past six years, she has managed the \$3 billion worth of programmes at the department's Office of Science.

Krebs says she is looking at a number of different career options. The department's Office of Science supports the majority of federally funded physics research in the United States, as well as extensive programmes in biology, chemistry, environmental science and other disciplines.

Eleven countries sign Framework deal

Munich Eleven central and eastern European countries scheduled to become full members of the European Union (EU) last week signed agreements with the European Commission that give them access to the union's fifth Framework Programme of research (FP5), on the basis of a financial contribution.

Researchers from these countries can now apply for funds from the 14.96 billion euro (US\$16 billion) programme, which runs from 1999 to 2002. In return, the eleven countries have agreed that researchers from EU member states can participate in the parts of their own research programmes which overlap with FP5's thematic programmes (see *Nature* **396**, 294; 1998).

Japan declines help for nuclear accident site

Munich Japan told the International Atomic Energy Agency (IAEA) last week that it would not be necessary "at this stage" for the agency to send an expert team to the nuclear accident site in Tokaimura, as the IAEA had offered.

Japanese nuclear authorities had informed the IAEA of the critical accident at the uranium-processing facility during the organization's general conference in Vienna,

which was attended by governmental delegates from 111 countries. But all information provided by Japan was "voluntary", according to the IAEA, rather than pursuant to the International Convention on Early Notification of a Nuclear Accident.

High-energy physics gets funding reprieve

Washington The US Congress has withdrawn, for the time being, its threat to withhold funding for research into future high-energy physics facilities, such as the Next Linear Collider, although it notes "serious concerns about the early cost projections of this planned facility".

A conference between the House of Representatives and the Senate to determine next year's funding for energy and water programmes rejected the Senate's plan to scale back support for the research (see *Nature* **400**, 390; 1999).

The conference also agreed to provide \$100 million for the construction of the Advanced Neutron Spallation Source at Oak Ridge Laboratory, Tennessee — half the amount requested by the Clinton administration — and boosted spending on fusion research by \$25 million to \$250 million.

Encyclopedic launch

The Nature Publishing Group and its sister publisher Macmillan Reference Ltd this week announced the online launch of the world's largest life-science reference work: *The Encyclopedia of Life Sciences*. Available in embryonic form, it can be accessed for a one-off fee of £125.

Subscribers can provide feedback as the encyclopedia evolves to the full 4,000 or more signed articles that will be available in print and online in 2001. The online publication will be regularly updated after completion.

Written by more than 2,000 leading scientists, it includes articles designed variously for students, postgraduates and fully fledged scientists. The scope of the encyclopedia ranges from biochemistry to evolution, and from bioinformatics to bioethics. Details of the publication, its advisory panel and authors, and ways of subscribing to the online embryonic form, can be found at www.els.net.

Correction.

Michael A'Hearn did not "head a recent review of the NOAO [National Optical Astronomy Observatories]", as we reported in our recent item on the Kitt Peak National Observatory (see *Nature* **401**, 199; 1999), but was a member of the committee to examine future directions for the NOAO.

Cash crisis threatens to close Mexico's leading university

Sir— Academic staff at the National Autonomous University of Mexico (UNAM), the most important university in the country, have been on strike for more than 160 days in protest at the government's plan to increase student fees. It has been a devastating experience.

Two hundred professors have been dismissed and there is a threat to close UNAM. It could be declared bankrupt so that the government would be able to cancel labour contracts without having to pay compensation. But it is vital that UNAM remains open. It educates more students than any other university in Mexico, many of them from other countries in Latin America. UNAM's 1,100 research scientists conduct 60% of the research done in Mexico.

Non-academic employees are also threatening to strike at the end of October unless they receive an absurd pay rise of 40%. This would spell the end of UNAM.

Mexico has an external debt of almost US\$200 billion, and the World Bank has imposed on the government tough economic measures which result in undesirable social and cultural effects. UNAM's financial requirements are now seen as a heavy burden on the government's budget and President Ernesto Zedillo seems intent on cutting such "excessive expenses". Yet all the economic problems of our country are a direct result of the actions of dishonest administrations over the years, who have been blind to the needs of science and technology.

UNAM is being seen as a business, which is an absurd mistake. Mexico has a small community of just 6,000 full-time researchers, 4,500 dedicated to natural sciences and engineering. Research receives far too little financial support.

UNAM has always played an outstanding role in science and it is imperative that this should be promoted and preserved above any other political consideration.

Alejandro Cuevas-Sosa

Centro de Prevención y Tratamiento de la Violencia Sexual e Intrafamiliar, San Lorenzo 203, Col. Del Valle 03100, Mexico, DF

How to bring collections data into the net

Sir— Electronic access to natural history collections data will enable biodiversity to be mapped with other parameters, improving science and environmental management. At least 10 major initiatives

desire networked access to these data¹. Studies of birds, important in environmental issues, may spearhead the development of networked databases². But discussions at this year's meeting of the American Ornithologists' Union in August offered a sobering assessment. Widespread participation by curatorial staff and researchers is required for a networked system to succeed, and several issues prevent this.

Perhaps 50% of specimens in North America have yet to be computerized, and proofreading lags far behind. Quality is an issue when releasing error-ridden data sets to point-and-click users who have little familiarity with the biases and errors in collections. Electronic users are unlikely to check data against specimens, the primary data source. Invoking 'downstream' quality control is not comforting, because peer review is a poor safety net rarely applied to reports and management plans.

Systematics collections are underfunded³, and the community disapproves of mining collections data while this is the case. Most data-sharing initiatives⁴ overlook this. The value of the data is recognized, so the enterprise producing them must be supported. The resource is also dated, on average reflecting environmental conditions of the 1920s⁴. Recent specimens and their data comprise the 'active zone', representing continuing activities to fill knowledge gaps in a changing world. Intellectual-property rights and institutional investments are concentrated here, as are the data most relevant to environmental management efforts. These data will not be widely accessible until bought and paid for. The solution is to renew support (in funding and permits) for fieldwork.

A two-step process towards networked databases may be acceptable. First, old data would be made available. Second, increased support would purchase access to the critical 'active zone'. Another viewpoint wants this support up front: commissioned biodiversity studies or subscriptions would generate new data and leverage access to the old.

The Freedom of Information Act causes concern in the United States⁵ because these databases are usually both public and private. But this act does uphold core values⁶. Data filters can protect vulnerable populations, intellectual-property rights, and rights to privacy.

The collections community is noted for sharing data, but leaving the tap on unminded may be an abrogation of responsibility. Progress on these issues, producing a model eliciting enthusiastic participation, is needed to fully access the wealth of data in natural history collections.

Kevin Winker

University of Alaska Museum,

907 Yukon Drive, Fairbanks, Alaska 99775, USA

1. *Association of Systematics Collections Newsletter* 27 (3–4), 1–23 (1999).
2. Kaiser, J. *Science* 284, 888 (1999).
3. PCAST *Teaming with Life* www.whitehouse.gov/WH/EOP/OSTP/NSTC/PCAST/pcast.html
4. Winker, K. *Conserv. Biol.* 10, 703–707 (1996).
5. Cohen, L. R. & Hahn, R. W. *Science* 285, 535–536 (1999).
6. Goldwater, W. H. *Science* 282, 1823 (1998).

Para-political force in attack on science

Sir— The Commentary by Benny Haerlin and Doug Parr, leading members of Greenpeace, is an insult to science and to the scientists who carry out their daily work on the basis of the scientific ethos (*Nature* 400, 499; 1999). Greenpeace is a para-political force, and one cannot expect that its activists will appreciate the inherent value system of the scientific enterprise. It is hardly qualified to advise on how to restore public trust in science.

Integration of academic research into industrial innovation does not necessarily imply that scientists lose their intellectual and moral independence. However, my colleagues and I agree with Greenpeace that academic research in the pursuit of truth requires freedom from commercial competition. I have enjoyed this freedom all my life.

In my experience as a science adviser to the government, dialogue with Greenpeace is difficult or impossible because its activists are not willing to accept scientific standards as the basis of communication. Haerlin and Parr also imply that the attitude of scientific institutions in demanding proof to justify an assertion or 'preventive action' is not acceptable to Greenpeace.

But this does not matter since the organization's goal is not scientific truth or neutral technology assessment but to sow mistrust in science and technology. Greenpeace has performed admirably in this regard. I agree with Haerlin and Parr that the prejudices firmly established in public opinion in Europe cannot, at present at least, be overcome by means of any science-based technology assessment. Over many years Greenpeace has contributed to the build-up of distrust between the scientific community and the public. It has undermined the credibility of scientists in assessing technologies.

I thought that *Nature* was devoted to the advancement of scientific thinking and practice. This hypocritical Commentary is clearly not intended to support the scientific community in its efforts to contribute to a higher degree of rationality and to a better world. The article is an attempt to further undermine the public prestige of science.

H. Mohr

University of Freiburg,

Schänzlestrasse 1, D-79104 Freiburg, Germany

Beyond 'substantial equivalence'

Showing that a genetically modified food is chemically similar to its natural counterpart is not adequate evidence that it is safe for human consumption.

Erik Millstone, Eric Brunner and Sue Mayer

Whenever official approval for the introduction of genetically modified (GM) foods has been given in Europe or the United States, regulatory committees have invoked the concept of 'substantial equivalence'. This means that if a GM food can be characterized as substantially equivalent to its 'natural' antecedent, it can be assumed to pose no new health risks and hence to be acceptable for commercial use. At first sight, the approach might seem plausible and attractively simple, but we believe that it is misguided, and should be abandoned in favour of one that includes biological, toxicological and immunological tests rather than merely chemical ones.

The concept of substantial equivalence has never been properly defined; the degree of difference between a natural food and its GM alternative before its 'substance' ceases to be acceptably 'equivalent' is not defined anywhere, nor has an exact definition been agreed by legislators. It is exactly this vagueness that makes the concept useful to industry but unacceptable to the consumer. Moreover, the reliance by policymakers on the concept of substantial equivalence acts as a barrier to further research into the possible risks of eating GM foods.

Acceptable daily intake

The concept of substantial equivalence emerged in response to the challenge confronting regulatory authorities in the early 1990s. Biotechnology companies had developed several GM foods and, to reassure their customers, wanted official approval for their introduction. But government statutes did not cover GM foods, nor did they provide the authority to regulate these innovations. Legislation could be amended, but that would not address the core problem of how to assess the risks. One obvious solution at that time would have been for legislators to have treated GM foods in the same way as novel chemical compounds, such as pharmaceuticals, pesticides and food additives, and to have required companies to conduct a range of toxicological tests, the evidence from which could be used to set 'acceptable daily intakes' (ADIs). Regulations could then have been introduced to ensure that ADIs are never, or rarely, exceeded.

From the point of view of the biotechnology industry, this approach would have had

two main drawbacks. First, companies did not want to have to conduct toxicological experiments, which would delay access to the marketplace by at least five years, and would add approximately US\$25 million per product to the cost of research and development. Second, by definition, using ADIs would have restricted the use of GM foods to a marginal role in the diet. An ADI is usually defined as one-hundredth of the highest dose shown to be harmless to laboratory animals. Thus, even if the animals show no adverse effects on a diet consisting exclusively of a test material, human intake would still be restricted to 1% of the human diet. The biotechnology companies want to market GM staples, such as grains, beans and potatoes, which individually might account for as much as 10% of the human diet, and collectively might provide more than half of a person's food intake.

The adoption of the concept of substantial equivalence by the governments of the industrialized countries signalled to the GM food industry that, as long as companies did not try to market GM foods that had a grossly different chemical composition from those of foods already on the market, their new GM products would be permitted without any safety or toxicological tests. The substantial-equivalence concept was also

intended to reassure consumers, but it is not clear that it has served, or can serve, that purpose. Although toxicological and biochemical tests, and their interpretation, are notoriously problematic and contested, and are slow and expensive, they can provide information vital to consumer protection¹.

Trying to have it both ways

The challenge of how to deal with the issue of risk from consuming GM foods was first confronted in 1990 at an international meeting, consisting of officials and industrialists but no consumer representatives, of the United Nations Food and Agriculture Organization (FAO) and the World Health Organization (WHO)². The FAO/WHO panel report makes intriguing reading, because what it fails to mention is as important as what is discussed. It does not use the term 'substantial equivalence' or mention ADIs. It implies that GM foods are in some important respects novel, but it then argues that they are not really novel at all — just marginal extensions of traditional techniques. These inconsistencies are inevitable, given that the industry wanted to argue both that GM foods were sufficiently novel to require new legislation — and a major overhaul of the rules governing intellectual property rights — to allow them to be patented,



Insubstantially equivalent? A GM food's toxicity cannot be predicted from its chemical composition.

yet not so novel that they could introduce new risks to public or environmental health.

The biotechnology companies wanted government regulators to help persuade consumers that their products were safe, yet they also wanted the regulatory hurdles to be set as low as possible. Governments wanted an approach to the regulation of GM foods that could be agreed internationally, and that would not inhibit the development of their domestic biotechnology companies. The FAO/WHO committee recommended, therefore, that GM foods should be treated by analogy with their non-GM antecedents, and evaluated primarily by comparing their composition with that of their natural antecedents, so that they could be presumed to be similarly acceptable. Only if there were glaring and important compositional differences might it be appropriate to require further tests, to be decided on a case-by-case basis.

Unfortunately, scientists are not yet able reliably to predict the biochemical or toxicological effects of a GM food from a knowledge of its chemical composition. For example, recent work on the genetics of commercial grape varieties shows that, despite detailed knowledge, going back for centuries, of the chemistry and flavour of grapes and wines, the relationship between the genetics of grapes and their flavour is not understood³. Similarly, the relationship between genetics, chemical composition and toxicological risk remains unknown. Relying on the concept of substantial equivalence is therefore merely wishful thinking: it is tantamount to pretending to have adequate grounds on which to judge whether or not products are safe.

The results of Arpad Pusztai's experiments with GM potatoes and their interpretation remain a matter of controversy⁴, but his starting hypothesis was that GM potatoes would be substantially equivalent to non-GM potatoes. Pusztai interpreted his still unpublished results as indicating that the GM potatoes exerted adverse biochemical and immunological effects, which could not have been predicted from what was known of their chemical composition. The experiments he conducted are not legally required and are therefore not routinely conducted before GM foods are introduced into the food chain.

'Substantial equivalence' ill-defined

The concept of substantial equivalence was first introduced in 1993 by the Organization for Economic Cooperation and Development (OECD)⁵, and was endorsed in 1996 by the FAO and WHO. Given the weight the concept has been required to carry, it is remarkable how ill-defined it remains, and how little attention has been devoted to it. The OECD document states:

"For foods and food components from organisms developed by the application of

modern biotechnology, the most practical approach to the determination is to consider whether they are substantially equivalent to analogous food product(s) if such exist. ... The concept of substantial equivalence embodies the idea that existing organisms used as foods, or as a source of food, can be used as the basis for comparison when assessing the safety of human consumption of a food or food component that has been modified or is new." That is the closest there has been to an official definition of substantial equivalence, but the definition is too vague to serve as a benchmark for public-health policy.

GM glyphosate-tolerant soya beans (GTSBs) illustrate how the concept has been used in practice. The chemical composition of GTSBs is, of course, different from all antecedent varieties, otherwise they would not be patentable, and would not withstand the application of the herbicide glyphosate. It is quite straightforward to distinguish, in a laboratory, the particular biochemical characteristics that make them different. GTSBs have, nonetheless, been deemed to be substantially equivalent to non-GM soya beans by assuming that the known genetic and biochemical differences are toxicologically insignificant, and by focusing instead on a restricted set of compositional variables, such as the amounts of protein, carbohydrate, vitamins and minerals, amino acids, fatty acids, fibre, ash, isoflavones and lecithins. GTSBs have been deemed to be substantially equivalent because sufficient similarities appear for those selected variables.

But this judgement is unreliable. Although we have known for about ten years that the application of glyphosate to soya beans significantly changes their chemical composition (for example, the level of phenolic compounds such as isoflavones⁶), the GTSBs on which the compositional tests were conducted were grown without the application of glyphosate⁷. This is despite the fact that commercial GTSB crops would always be treated with glyphosate to destroy surrounding weeds. The beans that were tested were, therefore, of a type that would never be consumed, while those that are being consumed were not evaluated. If the GTSBs had been treated with glyphosate before their composition was analysed, it would have been harder to sustain their claim to substantial equivalence. There is a debate in the research community on whether such changes to the chemical composition are desirable or undesirable, but it is an issue that remains unresolved, and which has been neglected by those who have deemed GTSBs and non-GM soya beans to be substantially equivalent.

Only one official organization has recognized some of the limitations of the concept of substantial equivalence. A Dutch government team has acknowledged that "compositional analysis ... as a screening method for

unintended effects ... of the genetic modification has its limitations ... in particular regarding unknown anti-nutrients and natural toxins", and it has given a lead by exploring some alternatives⁸.

The Dutch team accepts that comparisons of relatively crude compositional data provide a very weak screen against the introduction of novel genetic, biochemical, immunological or toxicological hazards, and they have suggested a finer-grained screen to test for differences in some of the relevant biological variables, such as DNA analysis, protein fingerprinting, secondary-metabolite profiling and *in vitro* toxicity testing. If such a finer screen revealed that a GM food contained a relevant novelty, the case for further studies would be far stronger, and those studies might benefit from having some clues as to which end-points should be investigated.

An anti-scientific test

Substantial equivalence is a pseudo-scientific concept because it is a commercial and political judgement masquerading as if it were scientific. It is, moreover, inherently anti-scientific because it was created primarily to provide an excuse for not requiring biochemical or toxicological tests. It therefore serves to discourage and inhibit potentially informative scientific research. The case of GTSBs shows, moreover, that the concept of substantial equivalence is being misapplied, even on its own terms, within the regulatory process.

If policymakers are to provide consumers with adequate protection, and genuinely to reassure them, then the concept of substantial equivalence will need to be abandoned, rather than merely supplemented. It should be replaced with a practical approach that would actively investigate the safety and toxicity of GM foods rather than merely taking them for granted, and which could give due consideration to public-health principles as well as to industrial interests. ■

Erik Millstone is at SPRU—Science and Technology Policy Research, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.

e-mail: e.p.millstone@sussex.ac.uk

Eric Brunner is in the Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK.

Sue Mayer is at GeneWatch UK, The Courtyard, Whitecross Road, Buxton, Derbyshire SK17 8NY, UK.

1. Food Additives and the Consumer Appendix I, 41–43 (European Commission, Brussels, 1980).
2. Strategies for Assessing the Safety of Foods Produced by Biotechnology (WHO, Geneva, 1991).
3. Bowers, J. *et al.* *Science* **285**, 1562–1565 (1999).
4. Masood, E. *Nature* **398**, 98 (1999).
5. Safety Evaluation of Foods Derived by Modern Biotechnology (OECD, Paris, 1993).
6. Lydon, J. & Duke, S. O. *Pestic. Sci.* **25**, 361–374 (1989).
7. Padgett, S. R. *et al.* *J. Nutr.* **126**, 702–716 (1996).
8. Kuiper, H. A. *et al.* Food Safety Evaluation of Genetically Modified Foods as a Basis for Market Introduction (Ministry of Economic Affairs, The Hague, 1998).

No ultimate truth in grand unification

That Theory of Everything is irrelevant to most physicists.

The Quest for Unity: The Adventure of Physics

by Etienne Klein and Marc Lachièze-Ray

(translated by Axel Reisinger)

Oxford University Press: 1999. 158 pp.

£17.99, \$29.68

George Ellis

There is a paradox at the heart of present-day physics that is not often remarked upon. It goes as follows. There is tremendous excitement among particle physicists, who are trying feverishly to develop a unified theory of all interactions and particles, and believe they are getting closer and closer to that goal. But the reality is that physics is becoming more and more fragmented. The unity that has been a great underlying theme in the development of physics is sharply contrasted with what is happening to the subject in practice.

Until recently, our best efforts at unifying gravity with the other forces were various superstring theories. These invoke a symmetry between fermions and bosons — previously understood to be quite different classes of particles — while giving a completely new perspective on the structure of space-time (now seen as 10- or 11-dimensional) and the nature of particles (which are now seen as vibration patterns of superstrings).

A major problem was the variety of such theories. The predominant effort now is to create a new theory ('M-theory') that will unite them all in a single structure, with symmetries (known as 'dualities') relating such apparently disparate theories.

The exact form of this supposed unification is not known, but weak-field limits are understood to some degree and may be the means of attaining the ultimate endpoint of theoretical physics — a Theory of Everything, in the sense of a theory unifying all fundamental forces into one. Success would complete the historical development that saw electricity and magnetism unified into electromagnetic theory, which was in turn unified with the theory of weak nuclear forces to give electro-weak theory. This may itself eventually be unified with the strong nuclear force to give a Grand Unified Theory. Although the probable structure of such a theory is broadly understood, such unification has not been fully achieved.

The paradox lies in the fact that this theoretical ferment, promising to transform the very foundations of the subject, is simply irrelevant to most physicists. It is of no major concern to, for example, solid-state physicists, plasma physicists or even those working in quantum optics and astrophysics. With the



The now defunct Superconducting Super-Collider: millions can be spent in the search for unity.

single exception of cosmology, the final form of this theory, if it is ever attained, will not affect what physicists do because the energies involved in its essential features are far too high to be relevant — indeed, so high that we may never be able to test the theory experimentally. So the vast majority of physicists continue to work in a pragmatic way in their separate fields, ignoring these somewhat esoteric endeavours of the theoretical physicists.

Even as the impetus for a fundamental unification of physics reaches fever pitch, the subject itself is becoming ever more fragmented into disparate subjects. Each has its own highly skilled specialists who, to a great degree, cannot even read the literature of the other specialist areas, or if they can, have neither the time nor the inclination to do so. Thus, this thrust towards unity — one of the most important driving forces of physics, which has led to the development of statistical physics, relativity theory and quantum theory — is in profound contrast to the way the subject is developing. And the importance of this becomes evident in disputes about funding, particularly for the enormously expensive particle accelerators such as the now defunct Superconducting Super-Collider. Particle physicists see these machines as vital to the most important experiments that can be done, and therefore of top priority, while many others consider them a waste of vast sums of public money that could be put to much better use within physics.

The Quest for Unity (ably translated from the original French) is a refreshing look at this tension between unity and diversity in physics, and places it in a useful historical perspective. The book touches on many issues of

interest in the philosophy of science, for example the relation between the eternally valid laws of universal application and the passage of time in changing physical systems possessing a unique identity; also, the way in which the abstract mathematical reasoning that underlies physics can form a foundation for venturing beyond the tested facts to new ways of understanding nature, which seems to be patterned in a mathematical way at a fundamental level.

The book clearly and concisely outlines the unifications already achieved by physics. It describes the core ideas of relativity and quantum theory, symmetries and conservation laws, supersymmetry and supergravity, and so on, and sets the scene for a critical review of the degree to which unification has been achieved. This leads to a discussion of the tension between realism and positivism, emphasizing that each discipline defines its approach by segregating off from all the rest what is of concern to itself; consequently, each sub-discipline has its own domain of description which largely ignores that of the others.

This isolationism is a prerequisite for successful physics, yet it is this differentiation of things that creates the tension with the view of the unity of all things. The book considers critically the underlying programme of reductionism which promises to reintegrate the different fields of study, but which has increasingly been at risk of making excessive claims that bring it into disrepute; for, in reality, this unification is mainly achieved in principle rather than in practice. A telling quote from Nietzsche harshly criticizes those who wish to replace the rich ambiguity of existence with highly idealized mathematical

models that are sometimes mistaken for reality itself.

This kind of reflection will become more important as physics moves closer to achieving its aims of fundamental unification while simultaneously diverging into ever more disjointed sub-disciplines. The authors conclude with the statement that only if physics renounces its pretensions to be a perfect reflection of reality will it continue to thrive: "Its greatest challenge is to resist the temptation to project the false self-image of a religion able to reveal the ultimate truth". Its practitioners will do well to heed this advice. The book provides a helpful platform for reflection on such issues. ■

George Ellis is in the Mathematics Department, University of Cape Town, Rondebosch 7700, Cape Town, South Africa.

Heritable traits of a grandfather

Erasmus Darwin: A Life of Unequalled Achievement

by Desmond King-Hele

Giles de la Mare: 1999. 422 pp. £24.99 (pbk)

W. F. Bynum

To most people who have heard of him, Erasmus Darwin was a successful doctor, bad poet and, most significantly, the grandfather of Charles Darwin. In this astonishing book, Desmond King-Hele seeks to reverse the judgement and argue that Charles should rather be remembered as Erasmus's grandson. Erasmus, King-Hele concludes, was much the brighter spark, a genius of rare qualities. In this book, 'Darwinian' refers to the grandfather and his legacy.

I doubt if the meaning of the adjective will be changed as a result of the book, but King-Hele's portrait of Dr Darwin is powerfully drawn. It is his sixth book, and second biography, devoted to his subject. The conclusions of his previous biography are unchanged, but the new availability of 170 family letters written by Darwin permits him to add much detail, including a valuable assessment of Darwin's relationship with his son (Charles's father), Robert Waring Darwin. Erasmus died before Charles was born, so the influence was through Robert Waring's recollections of his own father or through Erasmus's publications.

That Erasmus Darwin cut a substantial figure in his own lifetime cannot be doubted. Of massive size, with a pock-marked face and a severe stammer, he was also highly sexed and attractive to women. Following the death of his first wife, he sired two daughters by the servant looking after his young children. He then courted a beautiful, rich, married woman 16 years his junior, writing her love

Erasmus courting

Friend Boulton! Take these ingots fine
From rich Potosi's sparkling mine;
With your nice art a Tea-vase mould,
Your art, more valu'd than the gold!
With orient pearl, in letters white,
Around it, 'To the Fairest', write;
And where proud Radburn's turrets rise,
To bright Eliza send the prize.

poetry (see above) when he was not busy tending her medical needs. After her husband conveniently died, the 50-year-old doctor won out over his younger rivals and had a further seven children by her.

Darwin's interesting domestic life did not get in the way of his medical practice. He was the pre-eminent practitioner in the industrializing Midlands, travelling as much as 10,000 miles a year around Nottingham, Lichfield and Derby, where he settled after his second marriage. No wonder he was concerned with the design of carriages and the building of canals.

For most of his life, his reputation was simply that of a good doctor (and genial host). His early publications were produced anonymously, Darwin fearing that general knowledge that he was an inventor, poet, industrial entrepreneur and botanist would make patients less likely to consult him. Nor was he especially good at exploiting his mechanical ingenuity. His copying machine went unpatented, as did his ingenious improvements to the steering mechanism and springs of his carriage. His designs for a steam carriage or his speculations about mechanized flight were confined to his private commonplace-book or to correspondence with close friends. Among the latter were many of the pioneers of the Industrial Revolution, including James Watt, Matthew Boulton and Josiah Wedgwood. Making and keeping friends were two of his most endearing and enduring traits.

The necessity of earning his living through practising medicine thus rendered Darwin a behind-the-scenes figure in both the Industrial Revolution and the world of letters. Only in his last decade, the 1790s, did he venture to come into the limelight. First came his long, two-part scientific poem, *The Botanic Garden*. This was followed by a massive synthesis of medical knowledge (*Zoonomia*), further scientific poetry, a prose work on botany and scientific agriculture, and an essay on female education. This last was written for his two illegitimate daughters, whom he had set up as proprietresses of a boarding-school for young women. Altogether, he published more than a million words, not bad for a busy doctor chary of print.

Nor was the quality of his insight swamped by the quantity of words he scribbled. King-Hele reckons he contributed insight to no fewer than 86 areas of human endeavour and knowledge, ranging from adi-

abatic expansion to weather maps, from artesian wells to water closets. His ideas about evolution were based on an appreciation of the significance of the fossil record, the reality of biological extinction and the immense age of the Earth. He also remarked that the first law of organic nature was expressed in the words, "Eat or be eaten". Why, then, is the 'Darwin industry' confined almost exclusively to the grandson, especially when evolution and survival of the fittest are among those areas of grandfatherly insight?

There are a number of reasons. Many of his best ideas and inventions lay hidden in his commonplace-book until long after his death. Then, too, poetry is not the usual way to expound scientific discovery, at least not in the modern world. Even Darwin's reputation as a major poet was short-lived. Despite his undeniable influence on the Romantic poets, Darwin was soon relegated to the stilted and formal versifiers of the earlier age. His poetry is easier to admire than to respond to. Nor did his evolutionary ideas, deism and sympathy for the French Revolution endear him to the conservative thinkers and politicians who dominated British life after the Terror.

But if Darwin's historical fortunes have not been all he deserved, he has at least been fortunate in his modern biographer. Like Darwin himself, King-Hele is both a poet and a Fellow of the Royal Society, and his sympathy for his subject is total. Few scientific lives have ever been so carefully and thoughtfully examined. There are no final words in history, but this is a biography for which the word definitive can be aptly applied. I just wish I liked Darwin's poetry more. ■

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

An economist's use for sand

The Ostrich Factor: Our Population Myopia

by Garrett Hardin

Oxford University Press: 1999. 176 pp. £15.95, \$22

Paul Demeny

Readers familiar with Garrett Hardin's extensive works will pick up this slim volume confident of a re-encounter with his long-rehearsed themes on population growth and its dire effects. In this they will not be disappointed. As the ornithological title of the book suggests, the message this time is delivered with amplified polemical verve. Economists, we are told, are ostrich-like: they deny that a population problem exists. They disregard the limits on resources that nature imposes on man. They are unaware of the



Curbing growth: there has been a marked ...

power of human reproductive capacity, and have wrong ideas on such matters as free trade, multiculturalism and immigration (they tend to favour them all). Hardin's message to them is distilled in pithy admonitions, such as: "Thou shalt not transgress the carrying capacity"; "the maximum is not the optimum"; and "sustainable growth is an oxymoron".

Whatever the merits of the arguments in support of such aphorisms, they suffer here from brevity. Hardin's 1993 book, *Living Within Limits* (Oxford University Press), addressed the same topics with a combativeness tempered with documentary detail, which is in very short supply here. On the book's dust-jacket, the reader is promised "an uncompromising critique of mainstream economic thinking". Coming from an eminent biologist that should be a bonus. Disappointingly, Hardin's commentary displays a disconcerting lack of familiarity with economics — not a good basis for interdisciplinary dialogue.

Hardin's penchant for citing and interpreting past, often ancient, thinkers to shed light on contemporary issues is delightful. But economists, even those long defunct, do not seem to have figured much on his reading list. He complains, for instance, that they do not use the word 'environment'. Take, he says, John Stuart Mill's *Principles of Political Economy* (1848): the word is not in its index. But, arguably, no one in the past century and a half has written more eloquently about the desirable balance between man and nature than did, in Book IV, Chapter 6 of that work, Mill.

Today's "ostriches" are even less appreciated by Hardin. He sees them as possessed by growth mania — they are heedless maximizers. Yet economists' maximizing is based on a belief that what people maximize is utility subject to constraints: more of something

must always be balanced against the value of what has to be given up as its price. Economists also insist that voluntary exchanges between people are value-creating acts: they benefit both parties. But Hardin thinks they are a zero-sum game: what one acquires the other loses. Thus, he claims — in the face of massive evidence to the contrary — that free trade is a menace to material betterment.

Like many biologists, he sees man as just another animal species, and so underestimates the human potential for intelligent responses to environmental change, material scarcity and deficiencies in social arrangements. Economists do know that entities with a physical component cannot grow for ever. However, they recognize that there is no point in projecting exponential increase into the indefinite future because humans' economic calculations should provide corrective action to continuing population growth long before T. R. Malthus's pestilence and vice come into play.

A debate of the issue would require a close look at contemporary demographic trends, which Hardin fails to take. Examination of the evidence would reveal that the extraordinarily rapid expansion in the size of the human population during the past half-century was driven by a fall in death rates — by most evaluations, but not necessarily Hardin's, a positive development. It would also show that there has been a major and apparently accelerating downward adjustment of fertility rates in the past quarter-century. As a result, population growth rates have been falling. Standard projections see a levelling-off at 8 or 9 billion, plausibly followed by a declining population size.

Hardin offers a page of citations from assorted contemporary ostriches revealing their myopia by discounting the importance of limits on our natural resources and by expecting that living standards will continue to improve. It would be fair to say that the lack of alarm reflected in those citations is based on the perception that the declining curve of demographic growth is a natural outcome of the continuing structural transformations in the world economy. As the terms of the economic calculus guiding reproductive decisions shift, 'utility-maximizing' individuals will choose to restrict their fertility.

That the outcome of this process is desirable, however, is far from assured. The result of this interaction of utility-maximizing individuals may not deliver demographic patterns that the individual may consider best. As every member of society has an interest in the demographic behaviour of others, it may be possible to change the social rules of the game affecting such behaviour to the benefit of all. The merit of this book, and of Hardin's past work, is that it draws attention to this possibility and to the paucity of discussion surrounding the issue.

Ironically, future manifestations of such

market failures are more likely to have the opposite effect to the one that concerns Hardin, namely, that individual demographic choices may result in fertility levels that are well below the rate of replacement, and will produce a rapid decline in population and population ageing. But at present the reverse pattern can still be a problem. The economist's remedy (systemic and structural, through market-orientated development) and the complementary programmatic approach (better reproductive services, education, empowerment of women) are prescriptions that can command wide consensus. However, sometimes, and in some places, they may act too slowly. Hardin thinks this and calls for draconian solutions. An attempt to offer specific advice on the matter was indeed the original motive for this book. In past work, he says, "I was inhibited by unacknowledged taboos against taking a Darwinian approach to population. ... At the risk of coming a cropper, I approach the Malthusian barrier once more."

And what advice do we receive? "The first thing to do would be to cancel income deductions for the third child in a family (and beyond). If that is accepted ... we can think about stronger measures." — End of prescription. The image of the paterfamilias in, say, Rwanda, filling out his tax form and worrying about his deductions is too bizarre to contemplate. Perhaps another book is needed, approaching the Malthusian barrier once more. Or perhaps the absence of a better prescription simply shows that any "repressive legal measures" for reducing fertility come at a cost that is unacceptable to a democratic society. Muddling through with the known, if less than perfect, remedies is preferable. ■

Paul Demeny is at the Population Council, One Dag Hammarskjöld Plaza, New York, New York 10017, USA.



... decline in fertility rates in the past 25 years.

Science in culture

A year in Pangaea

Walking With Dinosaurs, a BBC Television series (first UK episode, 4 October)

Henry Gee

Makers of films about dinosaurs march in the footsteps of giants. There can be few who cannot remember watching, as awe-struck children, the doom-laden dinosaurs in Walt Disney's *Fantasia*, marching across the desert to the strains of Igor Stravinsky's *Rite of Spring*.

Then came *Jurassic Park*, Steven Spielberg's film (based on Michael Crichton's novel) in which the latest technology brought dinosaurs startlingly to life. *Jurassic Park*, said the publicists, was a film "65 million years in the making". The BBC's new series makes greater claims for itself. "Forget *Jurassic Park*," trumpets the hype. *Walking With Dinosaurs* is "the biggest thing on television in 200 million years".

In a series of six, half-hour episodes, *Walking With Dinosaurs* gives us snapshots of life from the Triassic (220 million years ago) to the end of the Cretaceous period (65 million years ago). Each episode follows, quite deliberately, the format of a standard wildlife documentary.

As a genre, such documentaries are remarkably formulaic. They are usually called *A Year in ...* [insert exotic locale of your choice], and start with a bloated, blood-red sunrise. They plot the lives of animals over the course of a year and contain lines such as "The herd of [insert herbivore name here] is nervous as a lone [append carnivore name of choice] approaches. The [carnivore] is desperately hungry. She hasn't eaten since last Tuesday fortnight and this is her last chance to feed her cubs/pups/chicks/kittens (delete as applicable) before the dry season. If she doesn't get a meal now, her [offspring name] will die — victims of the ceaseless, merciless Struggle for Existence." The credits roll over a bloated, blood-red sunset.

Judged from the first episode, at least, *Walking With Dinosaurs* follows this tradition, charting the lives of



various creatures over the course of what might be called *A Year in Pangaea*, weaving — in the cause of reanimation — known fact with various degrees of speculation.

In the first episode, for example, we meet a breeding pair of mammal-like cynodonts living in a burrow (for which evidence exists), laying eggs (a reasonable inference), suckling their young (rather more speculative) and pair-bonding for life (ditto). A baby cynodont is eaten by a marauding *Coelophysis* (a predatory dinosaur) and the cynodont couple, looking like demented, mutilated badgers, waddle disconsolately off (stage left) to make a new home.

As a wildlife 'docu-soap', it all works beautifully. What is worrying, though, is the mixture of fact and speculation melded into a seamless whole: this is fine for drama or science fiction, but I question whether it is entirely proper for something billed as a science programme. This is where comparisons with *Jurassic Park* become rather awkward.

In *Jurassic Park*, for example, the predator *Dilophosaurus* has bright warning coloration and the ability to spit venom. We could not possibly know or infer, from the fossil remains, that *Dilophosaurus* has any such attributes: but this is an example of precisely the point that *Jurassic Park* is trying to make. That is, if people insist on tampering with nature, such as by recreating dinosaurs from ancient DNA, they will be forced to contend with a variety of unguessable consequences.

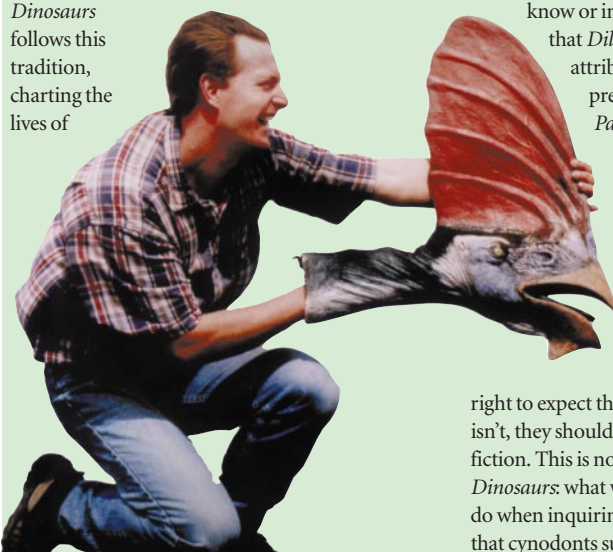
Things are different for a wildlife documentary, in which viewers should have the right to expect that everything is true; or, if it isn't, they should be able to tell fact from fiction. This is not the case for *Walking with Dinosaurs*: what will parents and teachers do when inquiring 12-year-olds insist that cynodonts suckled their young? It was

on the BBC, so it *must* be true.

But more ominous problems loom. In the cause of narrative, the programme discusses the evolutionary fates of creatures in very old-fashioned terms, talking of "the day of the dinosaurs" and "missing links" as if evolution were nothing more than a *scala naturae* animated by natural selection. This way of thinking of evolution — as the working out of preordained fate — was on its last legs in the 1920s and completely abolished by the advent of phylogenetic systematics in the 1960s and 1970s. Why, when palaeontology has come so far, do programme-makers still peddle concepts as antique as alchemy?

Leaving these problems aside, the script is so relentlessly monotonous that even the prodigious skills of the narrator — the actor Kenneth Branagh — cannot animate it. Everything in the Mesozoic world is seen as doom-laden: an animal must be constantly vigilant in case something bigger threatens to bite its head off. A sequence in which a small pterosaur takes a bath in a puddle could have been an excuse for some lightness in the gloom, but no: we are told that the benighted creature takes a terrible risk and doesn't know the grim end that fate has in store for it. This depressing message of doom, doom and more doom is emphasized by a score that sounds like one long, half-remembered excerpt from Gustav Holst's *Mars, Bringer of War*.

Which brings us back to *Fantasia*. Truly, a programme on dinosaurs, even one as ambitious as this, cannot escape its influences. A sequence in the first episode has a herd of *Placerias* (looking like mobile sofas with teeth) trudging across the desert — a scene guaranteed to bring *Fantasia* to mind. The genius of the Mouse (aided, in this case, by Leopold Stokowski) was to think of a Stravinsky soundtrack. In *Walking With Dinosaurs*, we have to put up with imitation upHolstery. *Walking With Dinosaurs* will be a hit, but as Noël Coward once said, one can only marvel at the potency of cheap music. Henry Gee is a senior editor at Nature.



Knowledge élites and class war

Would life be better if we left the difficult decisions to experts alone?

Sheila Jasanoff

For centuries, the gulf between the haves and the have-nots has been defined in material terms. The rich have more possessions, greater mobility, better nutrition and longer lives than the poor. While such disparities still separate rich nations from poor ones, within today's affluent societies class lines are a lot less easy to draw. Even average citizens have the means to own computers, fly in aeroplanes, eat food grown in far corners of the Earth and enjoy the miracles of modern medicine. At the same time, wealth seems insufficient to shield the rich from many risks of contemporary life, such as terrorism, environmental cancer and aeroplane accidents.

Knowledge, not money, has emerged as the most important instrument of social separation in technologically advanced societies. Those who possess scientific and technical expertise are spearheading the new industrial revolution in biotechnology and computers. Those who do not seem increasingly to be relegated to the status of passive consumers, their preferences shaped by the entrepreneurs of television and the Internet.

The redrawing of class lines through science and technology carries not only economic but also political consequences. Just as property ownership was once a condition for voting, so today there is a growing sentiment among the knowledge élite that public decision-making should be left to the experts who control specialized information. Gone is the idealistic, if sentimental, egalitarianism of the 1960s. From US calls for 'sound science' as the basis for regulating genetically modified foods to the explosion of expert advisory bodies worldwide, the knowledge meritocracy is everywhere asserting power.

One platform for the new class struggle between experts and non-experts is the US legal process. The recent vogue for letting judges screen expert testimony, and even appoint their own experts, bespeaks a growing impatience among professional élites with lay juries. This loss of confidence in the public's capacity to make sense of complex disputes bodes ill for the future of democracy. It also misreads the evidence about what the public knows and understands.

Are the public in developed countries hopelessly illiterate about science and technology, and should they therefore be kept at arm's length from the conduct of law and policy? Some would argue so. When a US jury acquitted O. J. Simpson of murder in October 1995, following the century's most widely



Trusting the experts: public acceptance of DNA tests has altered dramatically over a short period.

watched criminal trial, there was much hand-wringing in scientific circles about lay misunderstanding of scientific evidence. If the jury had only grasped the biological foundations of DNA fingerprinting, critics argued, it would not so cavalierly have acquitted Simpson. An educated appreciation of the technique's precision would have overwhelmed any residual doubts about police probity.

Ironically, just three years later, the US public unquestioningly accepted DNA evidence about the amorous proclivities of two of the nation's presidents. Historians had speculated for years that Thomas Jefferson, the revered third president of the United States, had fathered illegitimate children by his slave, Sally Hemings. Yet it was not until November 1998, when *Nature* published DNA test results from the Jefferson family and Hemings' descendants, that mainstream historians, journalists and the public accepted the conjecture as true.

Similarly, many who believed President William Jefferson Clinton's professions of sexual innocence in the early months of 1998 changed their minds on hearing about the DNA tests on Monica Lewinsky's famously unlauded blue dress. On this occasion, the US public seemed perfectly ready to accept the experts' assertion that it was not spinach dip that had caused the tell-tale stains. Despite the political character of independent counsel Kenneth Starr's investigation, a public wise to presidential weakness failed to express the kind of doubt that undermined the DNA evidence in the equally charged Simpson case.

Examples such as these should caution us against simplistic generalizations about how

the public regards scientific facts and claims, let alone against legal and administrative reforms that widen the gap between experts and non-experts. If the technological disasters of the late twentieth century — Bhopal, Chernobyl, the Challenger, environmental degradation — tell any coherent story, it is that expert assessments need to be tempered by broader visions. The risks of modernity are compounded by compartmentalizing technical knowledge and practical experience, separating lay from expert judgement, and mistaking the incomplete models of the technically literate for the sum total of reality. Would the Challenger have been launched, one wonders, if the engineers' doubts had been exposed to public scrutiny? And might the Green Revolution's environmental impacts have been lessened if farmers had played a more active role in designing the technology?

The strength of the common-law system historically has been to promote the integration of expert knowledge with lay perceptions of facts and values. This model of decision-making should be especially prized at a time when our sciences have made us sharply aware of the interconnectedness of things. As in current US debates about law reform, it may be tempting, in the short run, for knowledge élites to shake their heads over public ignorance and to avoid lay involvement in decisions affecting science and technology. But in the long run our hope lies in enhancing, not curtailing, the opportunities for conversation between science and society. ■
Sheila Jasanoff is at the John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA.

Condensates in a twist

Daniel S. Rokhsar

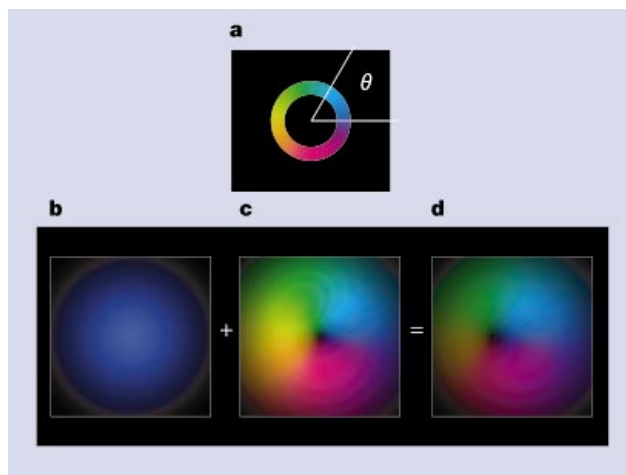
Studies of Bose–Einstein condensates continue to provide dramatic displays of quantum-mechanical behaviour. The long-sought goal of creating a quantized vortex in these unusual fluids has now been achieved.

The defining property of a fluid is a capacity for flow. At temperatures close to absolute zero, however, most materials freeze solid, and flow becomes impossible. Quantum mechanics predicts that the handful of systems that remain fluid under these extreme conditions should exhibit flow that is decidedly peculiar compared with our everyday experience. Groups at JILA and the University of Colorado, both in Boulder, Colorado, have taken the first steps in demonstrating these odd behaviours in a gas of ultracold atoms. On page 568 of this issue¹, Williams and Holland propose a novel method for inducing flow in a multicomponent quantum fluid; and last week, in *Physical Review Letters*, Cornell, Weimann and their colleagues² reported their implementation of this protocol to create a quantized vortex in a dilute Bose–Einstein condensate. These advances pave the way for in-depth studies of superfluidity in these systems.

Bose–Einstein condensates are quantum fluids in which many particles share the same quantum state, known as the ‘condensate’. For example, at absolute zero temperature one-tenth of the atoms in liquid ⁴He are in the condensate; the remaining atoms are kept out by the constant collisions within this dense liquid. Nearly 75 years ago, Bose and Einstein introduced the idea of a condensate by considering the opposite — and, at the time, purely hypothetical — case of a dilute gas. Such an idealized system was finally created four years ago by the JILA group³, who trapped thousands (and later, millions) of alkali atoms in a cloud tens of micrometres across, and cooled them to within a millionth of a degree above absolute zero. For these dilute systems collisions between atoms are so rare that essentially all the atoms can enter the same state at low temperatures, providing a textbook example of Bose–Einstein condensation. Over the past four years, dozens of laboratories around the world have prodded and probed condensates to examine their responses and test theoretical predictions. But a direct demonstration of the odd flow properties expected in a Bose–Einstein condensate has proved elusive.

According to quantum mechanics, the flow of a Bose–Einstein condensate is governed by variations in the condensate wavefunction’s phase, an angular variable θ that can be represented as a colour (Fig. 1a). Spatial variations in the phase determine the

Figure 1 Flow in a Bose–Einstein condensate. a, The phase angle θ can be represented as a colour selected from a colour wheel. Note that a 360° rotation of the phase samples all angles before returning the condensate wavefunction to its initial value; similarly, all colours of the wheel are encountered upon a full rotation, returning to the initial colour. b, A condensate at rest has uniform phase; here and in c and d the intensity of the colour represents the density of the condensate. c, In a unit vortex all colours are sampled as the centre is encircled. The density of the condensate vanishes at the centre because the phase becomes ill-defined there. d, Superimposing (b) and (c) leads to an asymmetric density profile as observed by Matthews *et al.*².



local flow velocity. The essential feature is that the wavefunction must be continuous, so that the phases sampled along any closed path must wrap around the colour wheel of Fig. 1a an integer number of times. In terms of flow, this translates into the constraint that the average velocity around any path multiplied by its circumference must be an integer multiple of the ‘quantum of circulation’ of h/m , where h is Planck’s constant and m is the mass of the particle.

We can now understand the unusual nature of flow in a Bose–Einstein condensate. At rest, the condensate phase is the same throughout (uniform colour, Fig. 1b), and the velocity is zero everywhere. The only way to introduce flow is to make sure that all phases occur at least once, which implies the presence of one or more regions around which all colours appear, as shown in the centre of Fig. 1c. This corresponds to a vortex — a pattern of circulating flow about a singular ‘core’ at which the condensate density must vanish. Near the core, the colour is changing rapidly with position, and the flow velocity is correspondingly large; near the periphery the colour changes more slowly, and the flow velocity is small. This kind of motion runs counter to our everyday experience for a rotating fluid — whereas the flow in a steadily rotating bucket of liquid water is slowest near the axis of rotation and fastest near the rim, the flow of a steadily rotating Bose–Einstein condensate is fastest at the

centre. Because the circulation must be an integer multiple of h/m , there is also a minimum allowed angular flow velocity below which no flow can occur.

Prior to the proposal of Williams and Holland¹, attempts to induce rotations in Bose–Einstein condensates had failed to generate vortices. Their new scheme exploits the possibility of simultaneously trapping otherwise identical atoms in two different internal ‘hyperfine’ states. (For example, although all ⁸⁷Rb atoms are identical in the sense that they each have 37 electrons orbiting a nucleus with 37 protons and 50 neutrons, there is flexibility in setting the relative spin of the various components within the atom.) So instead of a single condensate, two interpenetrating condensates of ⁸⁷Rb can be created. Each can be manipulated independently; furthermore, properly tuned microwaves can be used for coherent conversion of atoms from one hyperfine state to the other, intermixing the two condensates quantum mechanically. Williams and Holland take advantage of this extra flexibility by rotating the two condensates around each other during this interconversion. As proposed¹ and observed², the result is a double condensate with one condensate at rest (uniform colour, Fig. 1b) in the centre of the cloud and the other in a unit vortex state (all colours sampled once, Fig. 1c).

How can the quantum nature of flow in a Bose–Einstein condensate be observed?

NATURE

100 YEARS AGO

Many of our readers who are acquainted with Mr. Percy S. Pilcher, and others who have only heard of him through his great enterprise and keenness in constructing and using aerial machines, will be very sorry to hear that his accident on Saturday last has proved fatal, and that he died at 2.40 on Monday morning. Mr. Pilcher, during the last few years, had been making a considerable number of experiments with the object of constructing a soaring machine which would propel itself. The writer of this note was present at one of his trials in August 1897, at the time when he was at work in designing a small light engine for propelling his machine, and communicated to this journal an account (with illustrations from photographs) of his experiments on that occasion ... Like his forerunner Otto Lilienthal, Mr. Pilcher has come to the same sad end, and now his name must be added to that already long list of pioneers in aerial navigation. The experiments causing the fatality took place on Saturday last at Stanford Hall, the seat of Lord Brayne, near Market Harborough. We gather from the *Times* that after several ineffectual attempts to start, a signal was given about twenty minutes past four, and Mr. Pilcher rose slowly in the machine until he had travelled about 150 yards, and had risen to a height of 50 or 60 feet. Then a sharp gust of wind came and the tail of the apparatus snapped. Instantly the machine turned completely over and fell to the earth with a terrible thud, Mr. Pilcher being underneath the wreckage.

From *Nature* 5 October 1899.

50 YEARS AGO

A symposium on "Psychological Studies of the Quality of the Population" was held by Section J (Psychology) at the Newcastle meeting of the British Association, with the president of the section, Sir Godfrey Thomson, in the chair. Prof. P. E. Vernon (London), introducing the subject, outlined the well-known facts of the differential birth-rate not only between different social classes but also between families of different intelligence-level within the same class. Most authorities agree that this should lead to a decline of 1½ or more points of average intelligence quotient per generation, and that the numbers of very bright children be halved, of feeble-minded children doubled, before the end of the century.

From *Nature* 8 October 1949.

Box 1: Condensation of like minds

Further proof of the vitality of the field of Bose–Einstein condensation came at a meeting last month*, when experts in atomic physics, condensed-matter physics and quantum optics converged to discuss developments ranging from substrates for quantum computation to new types of trapped quantum fluids.

As discussed in the main text, Bose–Einstein condensation depends on the development of a coherent quantum phase across the fluid. Like any physical quantity, however, this phase is subject to thermal fluctuations, which are notoriously strong in low-dimensional materials and can alter or destroy condensation. Experiments that trap and cool gases of atomic hydrogen against a liquid helium surface have provided evidence of coherence in a dilute two-dimensional quantum fluid (T. Hijmans, Univ. Amsterdam; S. Jaakola, Univ. Turku).

Other developments foreshadow new directions for

trapped quantum systems. The essence of computation is the controlled transformation of input into output by the execution of a well-defined sequence of operations. Over the past few years it has been realized that if these operations are quantum mechanical in nature, dramatic advances in computational speed can, in principle, be achieved. Practical realizations of these schemes, however, have lagged behind theory. A new possibility is that regular lattices of cold atoms could be used for this purpose, and there are schemes for obtaining and manipulating such an array (P. Zoller, Univ. Innsbruck).

Bose–Einstein condensates are no longer the only quantum gases around. Although bosonic atoms such as ⁸⁷Rb can undergo condensation, others like the fermionic atom ⁴⁰K are forbidden from occupying the same quantum state, in the same way that electrons within atoms cannot occupy the same orbital and spin state. (Whether an atom is

a fermion or a boson depends on its total internal angular momentum.) In the first realization of a dilute atomic Fermi system⁸, nearly a million ⁴⁰K atoms were trapped and cooled to below microkelvin temperatures. By trapping the atoms in two internal states, and then cooling them, it has proved possible to overcome the fundamental quantum-mechanical limitations on the evaporative cooling of a single fermionic species (B. DeMarco and D. Jin, JILA and Univ. Colorado, Boulder). As predicted theoretically, the cloud contracted upon cooling until it reached a limiting size set by the Pauli exclusion principle. The search for pairing of atoms at still lower temperatures (analogous to the pairing of electrons in superconductors) promises to drive further developments in this new class of trapped quantum gases.

D. S. R.

*Bose–Einstein Condensation in Atomic Vapours, San Feliu de Guixols, Spain, 11–16 September 1999.

Matthews *et al.*² imaged the quantum phase of a vortex using an ingenious scheme that again takes advantage of the two-component nature of their condensate. They used an appropriately designed pulse of microwaves to simultaneously and coherently transfer atoms from the resting condensate to the vortex condensate. After the pulse, all of the atoms have the same internal state, and the resulting condensate is the sum of the two pre-pulse condensates (Fig. 1d), which corresponds to a displaced vortex⁴. This *in situ* superposition experiment elegantly demonstrates the 360-degree phase rotation of the vortex, a result that had been implied by the classic experiments of Vinen⁵ and Yarmchuk *et al.*⁶ in liquid helium, but never imaged so directly.

The generation of vortices in a Bose–Einstein condensate promises to illuminate the intimate details of the relationship between condensation and superfluidity, a link established by over 60 years of studies on liquid ⁴He. Superfluids are characterized by their ability to support dissipationless flow. In a conventional fluid the disorderly microscopic motions of the constituent particles are ultimately responsible for viscous drag; in a Bose–Einstein fluid this mechanism is suppressed by condensation. In the same

issue of *Physical Review Letters* as the JILA group's report², Ketterle and his collaborators at MIT described evidence⁷ for a critical velocity for dissipation in a Bose–Einstein condensate. (These and other developments in the field were presented at a lively conference in Spain last month — see Box 1.)

Superfluids can also support circulating currents that persist indefinitely, a phenomenon that can be understood in terms of the stability of vortex-like states of flow. Future observations of the relative motions of the core and periphery of a vortex, and the mechanism by which vortices enter and leave the condensate, will provide a deeper understanding of the link between superfluidity and condensation.

Daniel S. Rokhsar is in the Department of Physics, University of California, Berkeley, and the Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
e-mail: Rokhsar@physics.berkeley.edu

1. Williams, J. E. & Holland, M. J. *Nature* **401**, 568–572 (1999).
2. Matthews, M. R. *et al.* *Phys. Rev. Lett.* **83**, 2498–2501 (1999).
3. Anderson, M. H. *et al.* *Science* **269**, 198–201 (1995).
4. Butts, D. A. & Rokhsar, D. S. *Nature* **397**, 327–329 (1999).
5. Vinen, W. F. *Proc. R. Soc. Lond. A* **260**, 218–236 (1961).
6. Yarmchuk, E. J., Gordon, M. J. V. & Packard, R. E. *Phys. Rev. Lett.* **43**, 214–217 (1979).
7. Raman, C. *et al.* *Phys. Rev. Lett.* **83**, 2502–2505 (1999).
8. DeMarco, B. & Jin, D. S. *Science* **285**, 1703–1706 (1999).

Cell cycle

Fools rush in

Helen Piwnicka-Worms

The cell-division cycle is the means by which cells accurately duplicate themselves and then divide in two. In recent years a general model has emerged in which protein complexes consisting of cyclins and cyclin-dependent protein kinases (Cdks) regulate progression through this cycle¹. In addition, 'checkpoints' exist to monitor the integrity and replication status of the genetic material before cells commit either to replicate their DNA (in S phase) or to segregate it (during mitosis)². These checkpoints are signal-transduction pathways whose effectors interact with cyclin/Cdk complexes to block the cell cycle. A report by Chan *et al.*³ on page 616 of this issue helps us to understand how checkpoints do this.

Why the need to block the cell-division cycle? If DNA is damaged, a delay in the cell cycle facilitates repair, minimizing the replication and subsequent segregation of damaged DNA. Indeed, loss of checkpoint integrity is a hallmark of many human cancers⁴. Cells respond to DNA damage by stopping the cell cycle either at a stage known as G1, before DNA replication (via the G1

DNA-damage checkpoint), or just before mitosis by the G2 DNA-damage checkpoint.

The p53 tumour-suppressor protein is an essential component of the G1 DNA-damage checkpoint⁵. In response to damage, cells that lack p53 initiate a delay in the cell cycle at G2, but they neither delay entry into S phase nor maintain the arrest at G2^{6,7}. These properties have important clinical implications, because agents that disrupt the G2 DNA-damage checkpoint selectively potentiate the cytotoxic effects of DNA-damaging agents on cancer cells that lack p53⁸. Although the function of p53 in controlling the G1 checkpoint is fairly well understood at the molecular level, the contribution made by p53 to control of the G2 DNA-damage checkpoint is less clear. Chan *et al.*³ have now begun to address this.

The p53 protein activates transcription of other proteins, among them 14-3-3 σ (ref. 9). The seven human 14-3-3 proteins are highly conserved, phosphoserine-binding proteins involved in cellular proliferation, checkpoint control and apoptosis¹⁰. In epithelial cells, expression of 14-3-3 σ is induced in a p53-

dependent manner in response to DNA damage and, when overproduced, 14-3-3 σ blocks the cell cycle at G2⁹.

To test the contribution of 14-3-3 σ to G2 checkpoint control, Chan *et al.* disrupted the gene encoding 14-3-3 σ in a human colorectal cancer cell line containing functional p53. After DNA damage, these cells arrested in the G2 phase of the cell cycle, indicating that the G2 DNA-damage checkpoint was intact. But the cells could not maintain cell-cycle arrest, and they ultimately perished by non-apoptotic cell death. These findings indicate that expression of 14-3-3 σ is not required for cells to initiate the DNA damage-induced G2 delay, but that it is critical for cells to sustain — and, perhaps, to recover from — the delay.

Why does loss of 14-3-3 σ allow cells with DNA damage to escape arrest and enter mitosis prematurely? In human cells, entry into mitosis requires at least two events: activation of a mitosis-promoting protein called cdc2 by the cdc25C phosphatase; and accumulation of active cdc2 in the nucleus. In cells containing functional p53, the DNA-damage checkpoint seems to block both processes. First, cdc25C is prevented from activating cdc2. This is done by maintaining cdc25C in a phosphorylated form that binds various 14-3-3 proteins, preventing the cdc25C from accumulating in the nucleus^{11,12}, and activating cdc2 there. But, because cdc25C does not bind 14-3-3 σ (ref. 3), this pathway is intact in p53/14-3-3 σ -deficient cells. Second, cdc2 is exported to the cytoplasm. Normally, cdc2, in a complex with cyclin B1, is constantly shuttled between the nucleus and the cytoplasm (Fig. 1). Cyclin B1 contains a nuclear-export sequence that facilitates the efficient export of this complex into the cytoplasm¹³. Loss of 14-3-3 σ seems to interfere with this second regulatory pathway, allowing cdc2/cyclin B1 complexes to accumulate in the nucleus and, ultimately, bypass the G2 checkpoint.

Clearly, the molecular underpinnings of the p53/14-3-3 σ pathway deserve further attention. Although cdc2 co-immunoprecipitates with 14-3-3 σ , we do not yet know whether 14-3-3 σ binds directly to cdc2/cyclin B1 or to other proteins in the complex. We also need to know whether 14-3-3 σ anchors cdc2/cyclin B1 complexes in the cytoplasm, as Chan and colleagues propose³, or whether it regulates shuttling of these complexes between the nucleus and the cytoplasm. Furthermore, although loss of 14-3-3 σ results in the nuclear accumulation of cdc2, this alone is not expected to cause entry into mitosis — cdc2 must also be activated. But it is not clear how. The cdc25C regulatory pathway seems to be intact in cells that lack 14-3-3 σ . Moreover, the wee1 tyrosine kinase, which negatively regulates cdc2, can, presumably, inactivate nuclear pools of cdc2.

Chan and colleagues also report that wee1 is present in 14-3-3 σ immunoprecipi-

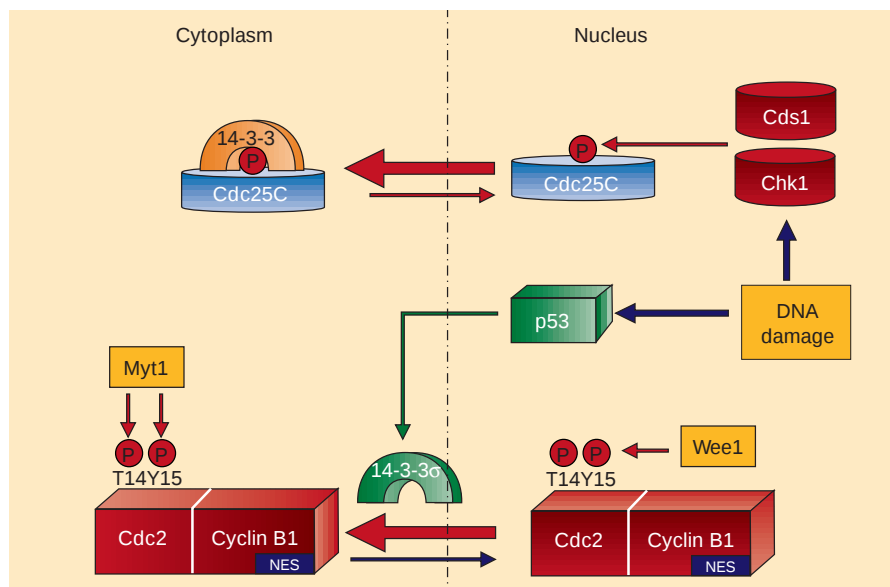


Figure 1 Cellular response to DNA damage. The cell-division cycle is blocked after DNA damage, allowing time for repair. After damage, cdc2 is maintained in an inactive state by phosphorylation on threonine 14 (T14) and tyrosine 15 (Y15). This is done by wee1, a nuclear kinase that phosphorylates cdc2 on Y15, and by Myt1, which is cytoplasmic and phosphorylates cdc2 on both T14 and Y15. Cdc2/cyclin B1 complexes continually shuttle between the nucleus and the cytoplasm, and a nuclear-export sequence (NES) in cyclin B1 prevents nuclear accumulation of cdc2/cyclin B1 complexes. Cdc25C, which activates cdc2/cyclin B1 complexes by dephosphorylating cdc2, also shuttles between the nucleus and the cytoplasm, and 14-3-3-binding proteins (although not 14-3-3 σ) contribute to the nuclear exclusion of cdc25C. DNA damage activates the nuclear Chk1 and Cds1 kinases, which maintain cdc25C in a 14-3-3-bound form. DNA damage stabilizes p53, leading to transcriptional activation of the 14-3-3 σ gene which, as Chan *et al.*³ have shown, contributes to the nuclear exclusion of cdc2/cyclin B1 complexes.

tates. So one possibility is that 14-3-3 σ facilitates interactions between cdc2 and wee1. In that case, loss of 14-3-3 σ might promote the nuclear accumulation of cdc2 and also disrupt the interactions between cdc2 and wee1. Together, these two events would be expected to generate active cdc2 in the nucleus and, in turn, to promote premature mitosis. Finally, future studies should be aimed at understanding why 14-3-3 σ -deficient cells choose mitotic catastrophe over adaptation as their response to DNA damage.

A model now emerges for epithelial cells in which distinct members of the 14-3-3 family target specific mitotic regulators to control various aspects of the G2 DNA-damage checkpoint (Fig. 1). Initiation of the checkpoint does not depend on p53, but it does require interactions between cdc25C and members of the 14-3-3 family. Maintenance of the checkpoint, however, depends on p53 and also involves interactions between cdc2, cyclin B1 and 14-3-3 σ . The 14-3-3 proteins use a common mechanism for initiating and maintaining the checkpoint — namely, exclusion of mitotic regulators from the nucleus. Components of the

p53/14-3-3 σ and cdc25C/14-3-3 pathways could, moreover, be therapeutic targets for drug development. In combination with established therapies that use DNA-damaging agents to kill tumour cells, such agents should enhance the anti-tumour effectiveness. Furthermore these agents are expected to be selective for cancer cells — because nonselective fools rush in where angels fear to tread. ■

Helen Piwnicka-Worms is at the Howard Hughes Medical Institute, Department of Cell Biology and Physiology, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110-1093, USA.

e-mail: hpiwnicka@cellbio.wustl.edu

1. Morgan, D. O. *Annu. Rev. Cell Dev. Biol.* **13**, 261–291 (1997).
2. Hartwell, L. H. & Weinert, T. A. *Science* **246**, 629–634 (1989).
3. Chan, T. A., Hermeking, H., Lengauer, C., Kinzler, K. W. & Vogelstein, B. *Nature* **401**, 616–620 (1999).
4. Hartwell, L. H. & Kastan, M. B. *Science* **266**, 1821–1828 (1994).
5. Kastan, M. B., Onyekwere, B., Sidransky, D., Vogelstein, B. & Craig, R. W. *Cancer Res.* **51**, 6304–6311 (1991).
6. Paules, R. S. *et al. Cancer Res.* **55**, 1763–1773 (1995).
7. Bunz, F. *et al. Science* **282**, 1497–1501 (1998).
8. Wang, O. *et al. J. Natl Cancer Inst.* **88**, 956–965 (1996).
9. Hermeking, H. *et al. Mol. Cell* **1**, 3–13 (1997).
10. Aitken, A. *Trends Cell Biol.* **6**, 341–347 (1996).
11. Peng, C.-Y. *et al. Science* **277**, 1501–1505 (1997).
12. Dalal, S. N. *et al. Mol. Cell Biol.* **19**, 4465–4479 (1999).
13. Yang, J. & Kornbluth, S. *Trends Cell Biol.* **9**, 201–210 (1999).

Environmental chemistry

End of the acid reign?

Alan Jenkins

From the onset of the Industrial Revolution in the eighteenth century right up to the mid-1970s, the emission of acidifying compounds to the atmosphere increased steadily. This trend, reflecting the growth of industrialization in Europe and North America, led to the release of sulphur and nitrogen compounds, largely from the burning of fossil fuels. These pollutants are transported in the atmosphere throughout the Northern Hemisphere and deposited as 'acid rain', causing acidification of soils and surface waters, with adverse effects on terrestrial and freshwater biota. In recent decades there have been considerable attempts to reduce the emission of problematic sulphur dioxide and nitrogen oxides, and it is encouraging to see new evidence that these expensive measures are beginning to take effect. For, on page 575 of this issue, Stoddard *et al.*¹ report that the chemistry of surface waters is showing signs that the acidification process is reversing.

The first legislation aimed at reducing sulphur emissions to combat the acidification problem was signed under the auspices of the United Nations Economic Commission for Europe (UN-ECE) Convention on Long-Range Transboundary Air Pollution in 1985 (also known as the First Sulphur Protocol). The signatory countries committed them-

selves to a 30% reduction in sulphur emissions (relative to the 1980 levels) by 1993. This was followed, in 1994, by the Second Sulphur Protocol, under which the signatories agreed to reduce emissions by about 60% by the year 2010, again relative to the 1980 levels.

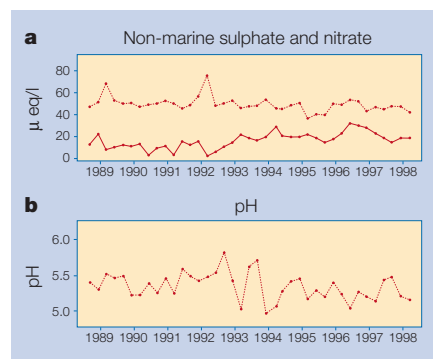


Figure 1 Effects of sulphate and nitrate deposition on surface waters. At Lochnagar in Northeast Scotland, one of the UK Acid Waters Monitoring Network sites, sulphate concentrations in the lake water have decreased over the past ten years in line with decreased deposition (a). Over the same period, nitrate concentrations have increased significantly (a), and, as a result, pH has apparently declined further (b). The reasons for the increased nitrate are not yet known.

These agreements have already prompted a significant reduction in sulphur dioxide emissions, although the link between emission and deposition is not linear, and, at any given location, depends on the proximity and direction of the major sources. The second protocol followed an effects-based strategy, using the critical-loads concept to direct emission cuts at those sources believed to cause the most environmental damage, and to afford the best protection to the most acid-sensitive ecosystems. (The critical load is defined as the quantity of pollution that a part of the environment can tolerate without harmful effects occurring, according to current knowledge.) In North America, Canada committed to the UN-ECE protocols, and, in 1991, the United States agreed to reduce national emissions of sulphur dioxide by 40% from 1980 levels by 2010.

Given that this legislation is relatively recent, that the technological measures needed for reduced emissions have long lead times, and that the deadline for achieving the target reductions is still ten years away, the chemical recovery observed by Stoddard *et al.*¹ across Europe and North America has been relatively rapid. The authors have done the first assessment of water chemistry data collected as part of the UN-ECE International Cooperative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes. The data are collected regularly by national agencies from around 200 sites across Europe and North America, and are collated within a database held at the Norwegian Institute for Water Research in Oslo. Stoddard and colleagues examined this database to assess chemical trends from 1980 to 1985, and to examine regional differences in these trends.

The observed recovery was somewhat expected, because evidence from small catchments covered with roofs^{2–4}, to remove sulphur deposition almost totally, showed a move towards chemical recovery almost immediately. But the authors did not see the same degree of chemical recovery everywhere. This result was also not unexpected. The severity of surface-water acidification at any site is a function of the total historical sulphur and nitrogen deposition, the soil buffering capability, and the rate at which base cations are supplied from primary weathering of bedrock. The degree and extent of recovery also depends on these factors, as well as on the reduction in sulphur deposition. Moreover, all of these factors are highly variable, even over relatively small areas, leading to difficulties in quantifying the 'regional' response to emission reduction through time.

All in all, the observations from long-term monitoring sites indicate that the international legislation is having a positive effect on the natural environment. But has the problem of surface-water acidification been

solved? At heavily acidified sites, where sulphur deposition is high, weathering rates are low, and soils are severely depleted in available (exchangeable) base cations, the chemical recovery will take a very long time — perhaps decades. Furthermore, we have no clear evidence of a biological recovery, and, indeed, know little about this process or its timescale.

Other factors may also offset the expected future recovery in response to reductions in sulphur deposition. A particular concern is the effect of nitrogen deposition (although this was recognized in 1988, with the adoption of the First Nitrogen Protocol under which signatory countries agreed to stabilize nitrogen emissions at their 1987 levels by 1994). There is evidence from experiments and from long-term records that terrestrial ecosystems are becoming less capable of retaining nitrogen. This has led to elevated concentrations of nitrate in soil and surface waters⁵. Because nitrate acts as a strong acid anion in aquatic systems, higher concentrations will promote increased acidity (Fig. 1). In August this year, following negotiations again led by UN-ECE, agreement was reached on a new protocol including emission ceilings on both sulphur and nitrogen.

Given the potential for changes in other ecosystem-driving variables to increase the acidification of surface waters or to offset the expected recovery from emission reductions alone, long-term monitoring is crucial. For example, changes in climate and land use may promote a change in the hydrological and temperature characteristics of a catchment system, causing increased decomposition of organic matter and possible increases in the concentrations of nitrate and sulphate in surface waters. So, only with an adequate understanding of ecosystem processes can we interpret the long-term trends and develop models that predict future responses to emission reductions or that optimize pollution-control strategies. Moreover, there are considerable gaps in our understanding, particularly with respect to the controls on nitrate leakage to surface waters, adsorption of sulphate by the soil, and the dynamics of weathering. We still have a lot to understand about the recovery process, and a long time to wait before surface waters are 'recovered'.

Alan Jenkins is at the Institute of Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK.
e-mail: jinx@wpo.nerc.ac.uk

1. Stoddard, J. L. *et al.* *Nature* **401**, 575–578 (1999).
2. Wright, R. F., Lotse, E. & Semb, A. C. *Can. J. Fish. Aquat. Sci.* **50**, 258–268 (1993).
3. Boxman, A. W., Van der Ven, P. J. M. & Roelofs, J. G. M. *Forest Ecol. Manage.* **101**, 155–164 (1998).
4. Moldan, F., Wright, R. F., Ferrier, R. C., Andersson, B. I. & Hultberg, H. in *Experimental Reversal of Acid Rain Effects: The Gårdsjön Roof Project* (eds Hultberg, H. & Skeffington, R.) 351–362 (Wiley, Chichester, 1998).
5. Wright, R. F. & Rasmussen, L. *Forest Ecol. Manage.* **101**, 1–363 (1998).

Immunology

Wavering on commitment

Suzanne Cory

It has the garb of an authentic immature B lymphocyte, but this progenitor blood cell actually has other agendas. If it is denied the transcription factor Pax5, and so cannot proceed any further down the path towards becoming a mature antibody-producing cell, this versatile cell can instead become a T lymphocyte or a natural killer cell, or assume the guise of a number of different bone-marrow-derived cell types: an antigen-presenting dendritic cell, a bacterium-engulfing macrophage, a bone-resorbing osteoclast or even a bactericidal granulocyte.

These unexpected observations, described in papers by Nutt *et al.*¹ and Rolink *et al.*² (pages 556 and 603 of this issue), challenge current ideas as to how immature cells become committed to producing specific types of progeny cells. The precedent of the MyoD family, which can entrain fibroblasts and some other cells to become muscle cells³, had led to the view that lineage commitment is the outcome of a master gene activating a suite of lineage-specific genes. The Pax5 results reveal an equally important job for such master genes: suppression of alternative choices of lineage.

The haematopoietic system, which generates all the various kinds of specialized cells in the blood, is superbly suited to studies on differentiation, thanks to a plethora of lineage-specific markers and excellent clonal culture systems. The patriarch of the system is a rare multipotential stem cell found in adult bone marrow. Through intermediate progenitor cells having diminishing options, it spawns cells of at least ten different lineages, including the various myeloid lineages, which control immediate responses to invading pathogens; the B and T lymphoid lineages, which confer specific long-term protection; and the erythroid lineage, which produces oxygen-carrying red blood cells.

Before birth, B cells are made in the liver and afterwards in the bone marrow. Commitment to B-cell production depends on the paired box transcription factor Pax5 (also known as B-cell-specific activator protein⁴) and two basic helix–loop–helix proteins, E2A and early B-cell factor (EBF) (see ref. 5). In the absence of either E2A or EBF, production stalls at the very early pro-B-cell stage, before any of the DNA recombination steps that create functional antibody genes. In the

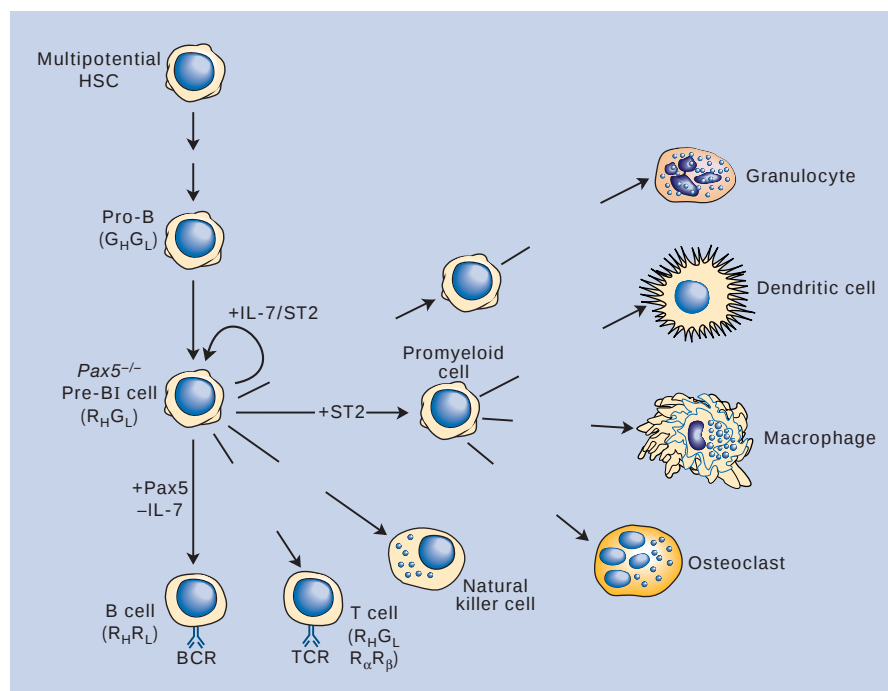


Figure 1 Developmental potential of Pax5-deficient B-lymphoid cells, as described by Nutt *et al.*¹ and Rolink *et al.*². In the absence of Pax5, B-cell development *in vivo* stalls at the pre-BI stage. If deprived of interleukin-7 (IL-7), Pax5-deficient pre-BI cells can take various differentiation pathways *in vitro*, the outcome depending on the available cytokine or cytokines provided in the medium and by ST2 stromal support cells. HSC, haematopoietic stem cell; SCF, stem-cell factor; CSF, colony-stimulating factor; BCR, B-cell receptor (antibody); TCR, T-cell receptor; G_H, G_L, R_H, R_L, R_α, R_β refer to germline and rearranged heavy- and light-chain immunoglobulin genes, and TCR α and β genes.

absence of Pax5, maturation in the bone marrow proceeds to the next stage, pre-BI, during which D_H and J_H heavy-chain gene elements recombine⁶, initiating the process that creates a unique antigen-binding site for the antibody. E2A and EBP cannot be sufficient for B-cell commitment because both are still expressed in cells deficient in Pax5 protein⁶.

B-lymphoid maturation can be recapitulated *in vitro*⁷. Pre-BI cells from bone marrow can be cultured for prolonged periods on support cells in the presence of interleukin-7 (IL-7), but, once removed from IL-7, they complete rearrangement of their immunoglobulin genes and differentiate into B cells bearing antibody receptors. As expected, Nutt *et al.*¹ found that pre-BI cells from Pax5-deficient mice were completely unable to differentiate into B cells when depleted of IL-7, unless Pax5 expression was restored.

The big surprise came serendipitously, when cultures that were left unattended for too long ran out of IL-7. Whereas the wild-type cells differentiated into mature B cells and died, the Pax5^{-/-} cells survived and many changed form, looking suspiciously like myeloid cells. Intrigued by this metamorphosis, Nutt *et al.*¹ started adding known myeloid cytokines to the IL-7-starved cultures. They found that the cells now exhibited a remarkable range of potentialities, as shown in Fig. 1. The cytokine M-CSF (macrophage colony-stimulating factor) induced terminally differentiated macrophages; GM-CSF (granulocyte-macrophage CSF) produced dendritic cells capable of activating T cells; Trance produced multinucleated osteoclasts; and G-CSF treatment following application of a cytokine cocktail produced granulocytes. The cells could also be persuaded to become natural killer cells, if cultured on support cells in IL-2. Even more strikingly, Rolink *et al.*² found that Pax5-deficient clones could fully reconstitute T-cell development in immunodeficient mice, and the resulting thymocytes were hallmarked by clonal D_H - J_H rearrangements indicative of their B-lymphoid origin, as well as having rearranged T-cell-receptor genes.

Significantly, single-cell sorting confirmed that individual Pax5^{-/-} pre-BI cells were multipotential — multiple subclones developed into macrophages, osteoclasts, dendritic cells or T cells, although production of natural killer cells and granulocytes was more variable. Clearly, Pax5-deficient pre-BI cells are not merely thwarted B-lymphoid cells. These cells could not save lethally irradiated mice, so not all haematopoietic options appear to be open to them. Nevertheless, it will be important to determine their full potentiality, both *in vitro* and in various *in vivo* settings. In particular, can they also be persuaded to become erythroid cells or platelet-producing megakaryocytes? Spleen-colony assays in irradiated mice

should shed some light on this question.

These studies^{1,2} confirm and extend previous evidence for the differentiation plasticity of certain early B-lymphoid cells (see ref. 8) and suggest that alternative fates are preserved in stem and progenitor cells by simultaneous low expression of master genes for several lineages⁹. Consistent with this view, Pax5^{-/-} pre-BI cells express genes from several lineage-affiliated programmes.

In normal B-lymphocyte development, the function of Pax5 thus appears to be just as much to suppress promiscuous transcription of genes specific for other haematopoietic lineages as to promote the expression of B-cell-specific genes. Complete suppression of other fates and full commitment to a B-cell future probably require a threshold level of Pax5 protein because, intriguingly, although both Pax5 alleles are expressed in immature B cells, only one allele is used in the earlier progenitors¹⁰.

The search is now on for the relevant target genes controlled by Pax5. Comparative screens of Pax5 mutant complementary DNA against that of normal pre-BI cells have identified several genes activated or repressed by Pax5, although to date none of them can alone account for the early developmental block¹¹.

Optics

Liquid versus photonic crystals

Eli Yablonovitch

A new paradigm has emerged in the past decade, in which the band-structure concepts of solid-state physics are applied to radio, microwave and optical waves. This has led to the invention of new forms of photonic crystal structures for controlling electromagnetic waves in three dimensions^{1,2}. These new structures are inspired by the three-dimensional geometry of both natural crystals and those that arise only in the human imagination. The latest step in this saga appeared in a paper in *Physical Review Letters*, in which Busch and John³ propose the first marriage of liquid crystals and photonic crystals. To appreciate this fully we have to recall the epic struggle to find an artificial crystal structure that would do for electromagnetic waves what a semiconductor crystal does for electron waves — a three-dimensional photonic bandgap that excludes a chosen range of frequencies.

The first proposals for a photonic bandgap^{4,5} called for face-centred-cubic (fcc) artificial crystals, because that geometry appeared to have the most promising interactions with electromagnetic waves. But it was surprisingly difficult to make a photonic bandgap and it was two years

The Pax5-deficient pre-BI cell lines are relatively robust and amenable to retroviral transduction and therefore should help to illuminate many aspects of haematopoietic differentiation. They are ideal test-tubes for screening for lineage-committing transcription factors, as well as for identifying their target genes. With inducible vectors, even the earliest steps in the developmental programmes should be accessible. The scientific fate of these fascinating cells, which flirt with so many destinies, seems assured. ■

Suzanne Cory is at The Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital, Victoria 3050, Australia.

e-mail: cory@WEHI.edu.au

1. Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. *Nature* **401**, 556–562 (1999).
2. Rolink, A. G., Nutt, S. L., Melchers, F. & Busslinger, M. *Nature* **401**, 603–606 (1999).
3. Weintraub, H. *Cell* **75**, 1241–1244 (1993).
4. Adams, B. *et al.* *Genes Dev.* **6**, 1589–1607 (1992).
5. Sigvardsson, M., O'Riordan, M. & Grosschedl, R. *Immunity* **7**, 25–36 (1997).
6. Nutt, S. L., Urbanek, P., Rolink, A. & Busslinger, M. *Genes Dev.* **11**, 476–491 (1997).
7. Rolink, A., Kudo, A., Karasuyama, H., Kikuchi, Y. & Melchers, F. *EMBO J.* **10**, 327–336 (1991).
8. Strasser, A., Elefanty, A. G., Harris, A. W. & Cory, S. *EMBO J.* **15**, 3823–3834 (1996).
9. Enver, T. & Greaves, M. *Cell* **94**, 9–12 (1998).
10. Nutt, S. L. *Nature Genet.* **21**, 390–395 (1999).
11. Nutt, S. L., Morrison, A. M., Dörfler, P., Rolink, A. & Busslinger, M. *EMBO J.* **17**, 2319–2333 (1998).

before an apparent success was reported⁶. That triumph was to be short lived. After some serious computations^{7,8}, a number of theorists agreed that it was actually a pseudo-gap masquerading as a real gap. This led to considerable concern, hand wringing and editorializing⁹.

The anxiety did not last long. Ho, Chan and Soukoulis¹⁰ discovered that a special

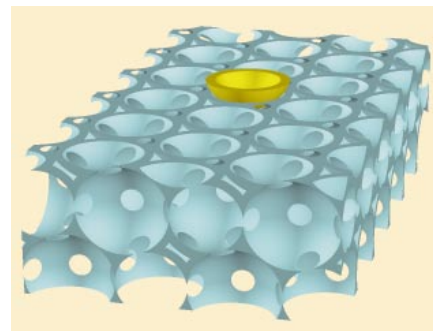


Figure 1 A marriage of liquid and photonic crystals. An inverse opal photonic crystal structure partially infiltrated with liquid crystal molecules, as proposed by Busch and John³. Electro-optic tuning can cause the photonic bandgap to wink in and out of existence

type of fcc packing, in which there are two atoms in each face-centred unit cell, possessed the long-sought photonic bandgaps. Those two atoms per cell are the famous diamond structure of engagement rings, hard industrial stone and silicon electronic chips — now reinvented for electromagnetic waves and Maxwell's equations. Diamond topologies quickly led to the first working photonic-bandgap crystal¹¹, albeit at inappropriately named 'microwave dimensions'. These structures are not microscopic at all. They are big and easy to make in a conventional machine shop, but their complex geometry is hard to fabricate at optical wavelengths^{12,13}.

Nature, of course, prefers to make simple fcc crystals. For example, SiO₂ spheres naturally self-assemble as fcc in opal, another jewel. As we saw, simple, easy-to-make fcc crystal structures had been proposed as photonic bandgaps^{4,5}, initially corroborated⁶, and finally rejected^{3,8,10}. But the tables were to turn again. The periodic structure of a crystal forces the frequencies to fit into a series of allowed bands. Everyone had been looking for a forbidden bandgap in fcc between the second and third bands. Gazing higher, between the eighth and ninth bands, Joe Haus and colleagues¹⁴ discovered a true forbidden gap in fcc; it was not a large one — only a few per cent of its central frequency — but it was a forbidden gap nonetheless.

The fcc geometry preferred by Maxwell's equations consists of spherical voids that are easily made by etching away spheres in a matrix of our choosing. Busch and John³ chose a silicon matrix, very appropriate to our times. They then added a new ingredient by calculating what would happen if the spherical voids of the photonic crystal were partially filled with liquid crystals (Fig. 1). We recognize liquid crystals as the electro-optic display molecules in portable computers. An electric field can rotate these molecules, thereby modulating the refractive index of the twisted nematic liquid crystal from $n = 1.4$ to $n = 1.6$, a fairly respectable change. Busch and John show that by applying an electric field to their hypothetical material a 2% photonic bandgap may be completely opened or closed. They give this new concept the rather glamorous name of a 'tunable electromagnetic vacuum'. Electromagnetic fluctuations are indeed being tuned in and out of existence, but they exist within tangible materials like silicon and liquid crystal molecules, rather than a vacuum.

In fact, as much as 40% of Busch and John's proposed photonic structure is a vacuum. To work correctly, the spherical voids must have their internal surfaces coated by the liquid-crystal molecules, so becoming half filled and leaving voids that may be difficult to achieve in practice. Nonetheless, Busch and John have linked forever the fields of liquid and photonic crystals. For example,

such materials could lead to light-emitting diode displays with pixels whose colours can be altered.

It is possible that the liquid/photonic crystal idea might bear fruit in structures even simpler than the three-dimensional fcc crystals. Two-dimensional photonic crystal semiconductor films have recently made the tiniest lasers, and probably the tiniest electromagnetic cavities, ever¹⁵. In conjunction with liquid crystals, they can be subject to extraordinarily delicate control, over lasing and other functions. Among possible applications, we may find photonic crystal lasers¹ in tiny optical integrated circuits, or we may see photonic crystals performing as simple white pigments or as radio wave structures in wireless information appliances, such as palmtop computers and sophisticated mobile phones².

Eli Yablonovitch is in the Department of Electrical

Engineering, University of California, Los Angeles, California 90095-1594, USA.

e-mail: eliy@ucla.edu

1. Scherer, A. et al. (eds) *IEEE J. Lightwave Technol.* (special issue on electromagnetic crystal structures) (in the press).
2. Scherer, A. et al. (eds) *IEEE Trans. Microwave Theory Techniques* (special issue on electromagnetic crystal structures) (in the press).
3. Busch, K. & John, S. *Phys. Rev. Lett.* **83**, 967–970 (1999).
4. Yablonovitch, E. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
5. John, S. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
6. Yablonovitch, E. & Gmitter, T. J. *Phys. Rev. Lett.* **63**, 1950–1953 (1989).
7. Leung, K. M. & Liu, Y. F. *Phys. Rev. Lett.* **65**, 2646–2649 (1990).
8. Zhang, Z. & Satpathy, S. *Phys. Rev. Lett.* **65**, 2650–2653 (1990).
9. Maddox, J. *Nature* **348**, 481 (1990).
10. Ho, K. M., Chan, C. T. & Soukoulis, C. M. *Phys. Rev. Lett.* **65**, 3152–3155 (1990).
11. Yablonovitch, E., Gmitter, T. J. & Leung, K. M. *Phys. Rev. Lett.* **67**, 2295–2298 (1991).
12. Lin, S.-Y., Fleming, J. G., Sigalas, M. M., Biswas, R. & Ho, K. M. *Phys. Rev. B* **59**, R15579–R15582 (1999).
13. Noda, S., Yamamoto, N., Kobayashi, H., Okano, M. & Tomoda, K. *Appl. Phys. Lett.* **75**, 905–907 (1999).
14. Sozuer, H. S., Haus, J. W. & Inguva, R. *Phys. Rev. B* **45**, 13962–13972 (1992).
15. Painter, O. et al. *Science* **284**, 1819–1821 (1999).

Photosynthesis

Stealing the limelight on the forest floor

By only opening their stomata — the pores in their leaves through which carbon dioxide enters and water leaves — at night, and thus achieving a temporary carbon dioxide fix using the enzyme phosphoenolpyruvate carboxylase, so-called CAM plants are able to minimize water loss. Not surprisingly, they are typically associated with arid environments. But a question mark hangs over certain bromeliads (members of the pineapple family) that grow in the deeply shaded understory of Central American rainforests. Such moist, low-light environments seem inappropriate for CAM plants, but the work of John Skillman *et al.* has shown that they occupy a niche in which high light tolerance and drought resistance have unexpected advantages (*Ecology* **80**, 1584–1593; 1999).

Most of the CAM plants in rainforests are epiphytes, growing perched on the branches of trees. Here, without access to groundwater, CAM metabolism helps them to survive occasional hot, dry spells. But there are also CAM plants growing in the shady understory. *Aechmea magdalenae*, pictured here, is a ground-dwelling bromeliad that grows in a range of habitats in



Central America, from well-drained, open locations to moist, shaded sites on the forest floor. Skillman and Winter previously found that *Aechmea* is capable of high photosynthetic rates under bright light (*Plant Physiol.* **113**, 441–450; 1997). But how can it compete in the shade — particularly as CAM plants are less efficient photosynthesizers than 'C₃' plants (the most widespread form of plant metabolism) where light levels are low?

One possible explanation is that *Aechmea* is particularly adept at exploiting the occasional patches of high-intensity sunlight ('sunflecks') that penetrate the dense canopy of the forest. Using chlorophyll fluorescence, which records the energy-trapping stage of photosynthesis, Skillman *et al.*

found that experimental artificial sunflecks produced a much stronger response in *Aechmea* than that found in C₃ plants from the same habitat, accompanied by a shorter lag phase for the plant to adjust to the new light conditions.

They also found that *Aechmea* allocates less energy to root formation, and grows more slowly, than associated C₃ plants. But it can gain an advantage over other plant species during the dry season. When soil water becomes scarce, the water conservation properties of the bromeliad, resulting from its tough leaves and its capacity to keep its stomata closed in the daytime, ensure survival and continued growth. *Aechmea* grows best during the dry season, whereas the C₃ plants grow best in the wet season. Although the forest floor is generally shady and moist, there is clearly a niche here for a plant that can take advantage of brief periods of high light and low water availability. Even among shade plants, there are some destined to grab the limelight.

Peter D. Moore

Peter D. Moore is in the Division of Life Sciences, King's College, Franklin-Wilkins Building, 150 Stamford Street, London SE1 8WA, UK.
e-mail: peter.moore@kcl.ac.uk

Bare bones of the cytoskeleton

Laura M. Machesky and John A. Cooper

Movement is a fundamental feature of living organisms and, on a molecular level, the mechanism of cell movement is highly complex. So it comes as an exciting surprise that Carlier and colleagues (page 613 of this issue¹) have dissected out a handful of proteins essential for reconstituting the motility of a bacterial propulsion system. The system concerned normally involves bacterial subversion of a host cell's own cytoskeletal machinery.

Cells adopt specialized shapes, change their shape and move around in response to various cues. In multicellular organisms, cell motility is essential for normal development and differentiation, as well as the response to disease. Cell shape and movement largely depend on actin, energy derived from the nucleotide ATP, and a host of other proteins associated with actin. Actin subunits spontaneously polymerize into filaments, which provide structural support and regulate the viscoelasticity and porosity of the cytoplasm; and polymerization of actin filaments is the driving force for extension of cell processes such as lamellipodia and filopodia. Adhesion and contraction also contribute to cell motility.

The pathogenic bacterium *Listeria monocytogenes* is often used for studying actin-based motility². Bacteria move within the

cytoplasm of host cells by recruiting cytoskeletal proteins to their surface. This results in actin polymerization at one pole of the bacterium, driving it forward. The set of components recruited by *Listeria* is presumed to include proteins involved in normal cell motility, so understanding how bacteria move should provide insights into how cells move. So far, analysis of the system has consisted largely of identifying components necessary for motility by depleting and inhibiting proteins in cells and cytoplasmic extracts.

Carlier and colleagues¹, however, have stripped the system down to determine its minimal requirements, and report the first reconstitution of bacterial motility from pure components. In addition to actin, only three other components are needed — the Arp2/3 complex, actin-depolymerizing factor (ADF, also called cofilin) and capping protein, all of which are well-known actin-binding proteins.

The bare bones of actin-based motility are now revealed, as depicted in Fig. 1. The Arp2/3 complex, composed of seven polypeptides including the actin-related proteins Arp2 and Arp3 (refs 3,4) nucleates actin polymerization^{5,6}. The complex binds the sides of actin filaments and nucleates the polymerization of new filaments, creating

a branching network⁶. The pointed (slow-growing) ends of the new filaments are bound to the Arp2/3 complex, and the barbed (fast-growing) ends are free. Actin subunits add to the free barbed ends, causing the filaments to grow. Over time, capping protein binds to the barbed ends, stopping their growth.

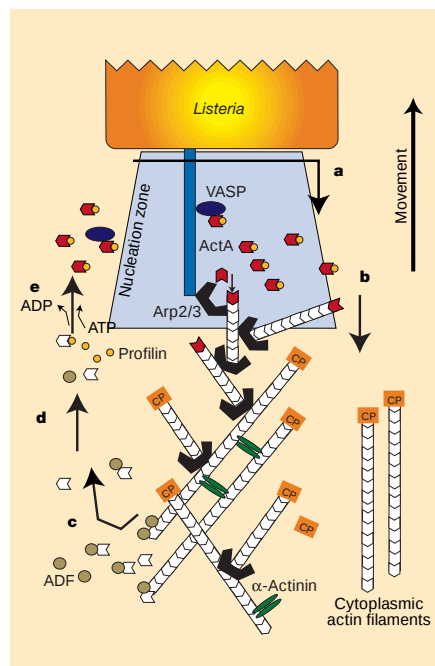
The need for capping protein seems paradoxical because it is expected to inhibit, not stimulate, actin polymerization. The paradox can be explained by a model⁷ in which the barbed ends of older filaments are capped. In consequence actin subunits add only to newly created free barbed ends, and polymerization is 'funnelled' to the new ends. In other words, capping protein restricts actin assembly to the zone nearest the bacterium.

The role of ADF is rather complicated and controversial. This protein stimulates the loss of actin subunits from pointed ends, which increases the pool of actin subunits that can be added to free barbed ends. It may also accelerate the rate of addition of subunits to free barbed ends⁸ (Fig. 1). The reason why increased filament turnover should be necessary for movement, instead of just stimulating it, is unclear.

For Arp2/3, capping protein and ADF, the dose-response curve for bacterial speed is bell-shaped (See Fig. 1b of the paper on page 614) — at low concentrations, speed increases with the concentration of the component; at high concentrations, speed decreases. High ADF may decrease speed by fragmenting filaments⁹ or inhibiting ATP-ADP exchange on monomers¹⁰. Excess capping protein may cap new barbed ends too quickly, and excess Arp2/3 may aggregate actin filaments or induce polymerization away from the bacterial surface. Notably, for each of the three components, the optimal concentrations for speed were slightly less than their concentrations in cytoplasm. Perhaps this was because the concentration of actin used in the assay was also less than that found in cytoplasm, due to technical considerations.

Carlier and colleagues¹ also report that three other proteins — profilin, VASP (vasodilator-stimulated phosphoprotein) and α -actinin — are important, but not necessary, for movement. Profilin, which sequesters actin monomers and increases bacterial speed, probably acts like capping protein in restricting (or funneling) actin polymerization to a limited zone near the bacterium. But its likely mechanism is different from that of capping protein; profilin appears to suppress spontaneous actin nucleation, so restricting nucleation to Arp2/3-induced free barbed ends¹¹. Other functions for profilin have been proposed, including increasing the rate of addition of subunits to barbed ends¹². But that effect required the presence of another protein that

Figure 1 Propulsion of *Listeria* inside a cell, as reconstituted by Carlier and colleagues¹. Essential components are the Arp2/3 complex; capping protein (CP); and actin-depolymerizing factor (ADF, also called cofilin). Stimulatory components are vasodilator-stimulated phosphoprotein (VASP); profilin; and α -actinin. a, The Arp2/3 complex and VASP bound to *Listeria* ActA promote assembly of actin filament branches. ATP-actin monomers (red chevrons) may be shuttled to the zone of nucleation by profilin bound to VASP/ActA or free in the cytoplasm. VASP may also interact with actin filaments to promote assembly. b, The branches eventually become capped at the barbed end by capping protein, and crosslinked by α -actinin, forming a stable scaffold known as the bacterial actin tail. Other actin filaments in the cytoplasm, away from the zone of nucleation, are probably likewise capped. c, At the pointed ends of filaments, ADF enhances disassembly of monomers from the filaments. d, Profilin will compete with ADF for binding to ADP-bound actin monomers (white chevrons), which will promote the exchange of ADP for ATP to re-charge the monomers for polymerization onto new barbed ends. e, Profilin-ATP-actin complexes will then be recycled for use in the nucleation zone. Profilin also prevents spontaneous assembly of ATP-bound actin monomers into nuclei in the cytoplasm outside the nucleation zone.



sequesters actin monomers, thymosin- β 4, which was not included in this study.

Profilin may also bind to VASP, the second stimulatory component examined by Carlier and colleagues. VASP interacts with the bacterial surface protein ActA, and may 'shuttle' profilin-actin subunits to the new barbed ends¹³. However, profilin mutants that do not bind VASP are able to stimulate motility¹⁴, and VASP without profilin has a stimulatory effect on its own, perhaps by directly interacting with actin filaments¹⁵.

Finally, α -actinin, which crosslinks actin filaments, is known to be necessary for bacterial movement in living cells¹⁶. Carlier and co-workers found that it did not influence the speed of bacterial movement but did affect the morphology of the tails — without α -actinin, tails were splayed, not compact, because filaments were not crosslinked. This crosslinking function may be essential *in vivo*, where the forces necessary for movement should be greater than those in Carlier and colleagues' purified *in vitro* system.

Other components, not tested here, may also be necessary or stimulatory for bacterial motility in cells. The requirements for motility *in vivo* may be different or more demanding than those in the reconstitution system. For example, the need for profilin may be greater if thymosin- β 4 acts to buffer the supply of actin monomers.

As well as providing a wonderfully simple model for actin-based motility, Carlier and colleagues' landmark study provides a new motility assay to analyse components and

regulators of the actin cytoskeleton. It complements other methods for measuring actin polymerization and network formation, and greatly increases our ability to analyse actin assembly *in vitro*.

Laura M. Machesky is in the Division of Molecular Cell Biology, School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK.

e-mail: L.M.Machesky@bham.ac.uk

John A. Cooper is in the Department of Cell Biology, Washington University, St Louis, Missouri 63110, USA.

e-mail: jcooper@cellbio.wustl.edu

- Loisel, T. P., Boujemaa, R., Pantaloni, D. & Carlier, M. F. *Nature* **401**, 613–616 (1999).
- Dramsi, S. & Cossart, P. *Annu. Rev. Cell Dev. Biol.* **14**, 137–166 (1998).
- Machesky, L. M., Atkinson, S. J., Ampe, C., Vandekerckhove, J. & Pollard, T. D. *J. Cell Biol.* **127**, 107–115 (1994).
- Welch, M. D., DePace, A. H., Verma, S., Iwamatsu, A. & Mitchison, T. J. *J. Cell Biol.* **138**, 375–384 (1997).
- Welch, M. D., Rosenblatt, J., Skoble, J., Portnoy, D. A. & Mitchison, T. J. *Science* **281**, 105–108 (1998).
- Mullins, R. D., Heuser, J. A. & Pollard, T. D. *Proc. Natl Acad. Sci. USA* **95**, 6181–6186 (1998).
- Carlier, M.-F. & Pantaloni, D. *J. Mol. Biol.* **269**, 459–467 (1997).
- Carlier, M.-F. *et al. J. Cell Biol.* **136**, 1307–1322 (1997).
- Maciver, S. K., Pope, B. J., Whytock, S. & Weeds, A. G. *Eur. J. Biochem.* **256**, 388–397 (1998).
- Blanchoin, L. & Pollard, T. D. *J. Biol. Chem.* **274**, 15538–15546 (1999).
- Machesky, L. M. *et al. Proc. Natl Acad. Sci. USA* **96**, 3739–3744 (1999).
- Pantaloni, D. & Carlier, M. F. *Cell* **75**, 1007–1014 (1993).
- Chakraborty, T. *et al. EMBO J.* **14**, 1314–1321 (1995).
- Egile, C. *et al. J. Cell Biol.* **146**, 1–14 (1999).
- Laurent, V. *et al. J. Cell Biol.* **144**, 1245–1258 (1999).
- Dold, F. G., Sanger, J. M. & Sanger, J. W. *Cell. Motil. Cytoskeleton* **28**, 97–107 (1994).

Plant defence

Long view from a high plateau

Jeff Dangl

Scientific endeavour tends towards a punctuated equilibrium — slow periods during which systems and tools are developed, followed by bursts of new knowledge. For the molecular understanding of plant–pathogen interactions there have been three recent explosions. The first was the isolation of plant disease-resistance (*R*) genes (Fig. 1); second was the ability to isolate mutants in 'model' plants such as *Arabidopsis thaliana* and tomato; and third was the (initially improbable) finding that bacterial pathogens of both plants and animals rely on a conserved delivery system to ferry the effectors of disease into their hosts. Experiments and achievements deriving from these three breakthroughs were reported at two meetings earlier this year*.

One current debate swirls around how the diversity of the *R* genes evolves and is maintained. This debate reminds me of (and draws from) discussions that followed the isolation of major histocompatibility complex genes in the early 1980s — full of structures and sequence comparisons, with enlightening forays into population genetics and molecular evolution.

Most *R* proteins contain leucine-rich repeats (LRRs), and there is overwhelming evidence that solvent-exposed surfaces of these repeats are subject to diversifying selection^{1,2}. Selection acts on point mutations and on short tracts of DNA exchanged between chromosomes by recombination and, probably, gene conversion. Many *R* genes are found in linked clusters on the chromosomes. At the *R* cluster of each parental chromosome, those *R*-gene sequences (or 'haplotypes') that derived from a common ancestor (orthologues)

seem to be more related than those that originated by duplications (paralogues; Richard Michelmore, Univ. California, Davis). If so, *R*-gene divergence is an ancient event³.

A related and equally compelling argument is that architecture of the *R*-gene cluster determines the recombinational outcome and influences subsequent diversity (Jonathan Jones, Sainsbury Laboratory, Norwich). Two haplotypes at a region known as *RPP5* in *Arabidopsis* are scrambled by rearrangements, so it is impossible to define orthologues here. Recombination between these two haplotypes is suppressed. The extraordinary divergence in the *R*-gene cluster is backlit by considerable homology between the DNA sequences that flank *RPP5*. By contrast, the tomato *Cf-9* and *Cf-4* genes are orthologues embedded in unique, but fairly linear, haplotypes, and evidence for recombination between these two genes can, occasionally, be found.

When combined, these results (Michelmore; Jones) indicate that, at the *R* clusters examined, unequal recombination events can influence the evolution of a particular stretch of DNA, but that they contribute little to its diversification. Evolution of an *R*-gene cluster can also be influenced by chromatin dynamics, a view supported by molecular analyses of the flax *L* and *M* regions. When sequences from 11 of the 13 alleles at *L* were compared, diversity was found to be generated by deletion and expansion of LRRs. This region is also subject to diversifying selection (Peter Dodds, CSIRO, Canberra). So, similar events occur at simple and complex *R* genes.

Diversity aside, what is the function of the *R* proteins? The simplest idea is that they act as receptors for ligands encoded by the *avr* genes, but this has been difficult to

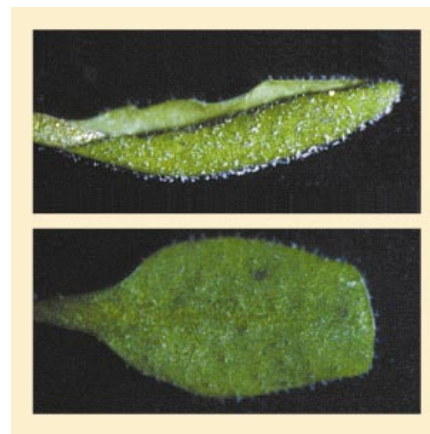


Figure 1 Downy mildew disease of many plants is caused by obligate oomycete parasites. *Peronospora parasitica* infects one *Arabidopsis thaliana* accession (top), but a second *Arabidopsis* inbred line is resistant (bottom) owing to the action of a single disease-resistance (*R*) gene.

* Attack and Defence. The Thirteenth John Innes Symposium 20–23 July, Norwich, UK; and Ninth International Congress of Molecular Plant–Microbe Interactions 25–30 July, Amsterdam, The Netherlands.

demonstrate. An LRR domain encoded by the rice *Pi-ta* gene has been shown to bind its cognate Avr protein, though, and this interaction can be terminated by exchanging a single amino acid on the carboxy-terminal side of the final LRR (Barbara Valent, DuPont, Wilmington). Interactions between R and Avr might require a third (or even a fourth) partner, and, although sequence variability in the LRRs determines specificity, a broad structural signature might be derived from an R-Avr complex containing other proteins. For example, a yeast three-hybrid screen was used to isolate partners for the AvrPto-Pto pair (Greg Martin, Cornell Univ.). Pto is a rather exceptional R-gene product — it has no LRRs and, although it binds AvrPto, it also requires a more catholic LRR-containing protein called Prf in order to function.

How can we identify more genes required for R function? Genetic screens for loss of resistance pick out only a few genes owing to redundancy or lethality. For instance, the barley Rar1 protein is required for cell death after recognition of a pathogen (Paul Schulze-Lefert, Sainsbury Laboratory), and its homologue in the nematode *Caenorhabditis elegans* is also involved in germ-cell death. Hence, what Schulze-Lefert terms “transgenomics” can identify conserved functions across kingdoms. Yet despite the difficulties, the spectrum of *Arabidopsis* mutants with effects on disease resistance grows continuously⁴. In conditional *avr*-expression screens of several hundred thousand progeny, for example, we have identified rare mutations of new genes required for *RPM1* function (Pablo Tornero and myself).

There is also redundancy in the steps that occur after the R protein has recognized its ligand. When parsley cells are exposed to a fungal protein that initiates defence responses, five mitogen-activated protein (MAP) kinases are activated (Dierk Scheel, Inst. Plant Biochem., Halle). All tested isoforms seem to activate defence genes in parallel. And, using the powerful ‘fast-forward genetics’ method, a calcium-dependent protein kinase has been identified upstream of other, probably redundant, MAP kinases that transduce the function of the *Cf-9* and *Cf-4* R genes (Tina Romeis, Sainsbury Laboratory).

The technique of fast-forward genetics was developed by David Baulcombe and collaborators (Sainsbury Laboratory), and is called virus-induced gene silencing (VIGS)⁵. This method allows the expression of any gene to be silenced, and it is being used to identify the genes required for a variety of R-gene functions. In principle, VIGS can ascribe function to genes with lethal loss-of-function effects, and it can be used to silence several (possibly redundant) members of one gene family if

they are more than 80% homologous. With the development of the tobacco rattle virus this method should soon be applicable in *Arabidopsis*.

Another question is how bacterial pathogens deliver the effectors of disease to plants and trigger the defence responses in the first place. The highly conserved type-III delivery system is encoded by the *hrp/hrc* cluster, expression of which is induced by contact with the host (Christian Boucher, CNRS-INRA). One border of the *hrp/hrc* pathogenicity island is flanked by a variable domain that can encode different effector proteins (Alan Collmer, Cornell Univ.). When expressed inside plant cells, these can trigger R-gene-dependent responses — that is, they behave like Avr proteins. They can also contribute to virulence in hosts that lack the appropriate R gene. Plant pathogens maximize their potential virulence by ‘collecting’ type-III effectors and deploying them blindly into potential host cells. If none of these effectors is recognized by a host R-gene function then they may contribute to virulence.

A newly discovered pathogenicity island encodes at least five type-III effectors, four of which contribute to virulence (John Mansfield, Wye College)⁶. These are linked in a sea of insertion sequences and transposons, suggesting that many genetic events have created a pile-up of potential virulence functions. We have found that the host response can, in fact, initiate the excision of a pathogenicity island from the pathogen’s chromosome. We are currently exploring whether this plasmid can be horizontally transferred to a new bacterial host. So the weaponry of bacterial pathogens can respond rapidly to the host’s defences.

These lines of enquiry will slow as the next wave of challenges form, challenges that are increasingly cell-biological and biochemical in nature. The effects of genomics are beginning to be felt, and, along with the further refinement of genetic screens, will provide fodder for the next evolutionary steps in this arena. The development of new tools and systems will pique anticipation of the next punctuation in the equilibrium. Like evolution, we will stumble down several blind alleys — but the way forward will inevitably be found. ■

Jeff Dangl is in the Department of Biology and Curriculum in Genetics, University of North Carolina, Chapel Hill, North Carolina 27599-3280, USA.

e-mail: dangl@email.unc.edu

1. Parniske, M. *et al.* *Cell* **91**, 821–832 (1997).
2. Michelmore, R. W. & Meyers, R. C. *Genome Res.* **8**, 1113–1130 (1998).
3. Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. *Nature* **400**, 667–671 (1999).
4. McNellis, T. W. *et al.* *Plant J.* **14**, 247–258 (1998).
5. Angell, S. M. & Baulcombe, D. C. *EMBO J.* **16**, 3675–3684 (1997).
6. Jackson, R. W. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 10875–10880 (1999).

Daedalus

Instant diamond

A diamond is a single crystal, a vast regular lattice of carbon atoms. But graphite can also be obtained in single-crystal form, both as a natural mineral and by slow annealing of polycrystalline graphite. Daedalus now proposes to flip one crystal form into the other.

The transition from graphite to diamond is fast and martensitic, taking only a few nanoseconds. A sudden pressure-pulse, compressing a perfect graphite crystal above the transition pressure while heating it above the transition temperature, should convert it to diamond. Daedalus wants to do this by explosive forming.

The explosive would be laid against a carefully shaped piece of single-crystal graphite, and detonated. The shock-wave traversing the graphite would collapse its lattice into the denser diamond one. The geometry of the process needs careful study. Each carbon atom must be projected into the graphite structure ahead of it, in just the direction to take up its proper place in the growing diamond lattice. Daedalus reckons that the explosion front needs to be normal to a main bonding direction in the graphite planes, and at a dihedral angle to them of about 22°. The transition should then be handed on through the graphite lattice as a self-replicating molecular mechanism.

The explosive formation of bulk diamond from single-crystal graphite will send a violent shock through the jewellery trade. But Daedalus goes further. Graphite reacts with many substances — alkali metals, halogens and ions of various kinds — which get between the layers of carbon atoms. The explosive transformation of such lamellar graphite compounds would give doped diamond.

Some natural diamonds are doped with nitrogen or other elements; they are often coloured. But solid-state and semiconductor physicists should leap at the chance to dope diamond with controlled amounts of other well-chosen impurities. Pure diamond is almost a perfect insulator. By contrast, doped diamond could be a splendid new semiconductor. Not only might it work at unprecedented temperatures and huge voltages. It could be the basis of many new logical, piezoelectric, thermo- and photo-voltaic devices. In particular, the transparency of diamond should make it ideal for LEDs and solid-state lasers. These cheap, mass-produced products would allow even ordinary folk to flaunt real diamond jewellery with a truly brilliant sparkle.

David Jones

Arsenic poisoning in the Ganges delta

The natural contamination of drinking water by arsenic needs to be urgently addressed.

The pollution by naturally occurring arsenic of alluvial Ganges aquifers, which are used for the public water supply in Bangladesh and West Bengal, has been discussed by Nickson *et al.*¹. We agree with their main conclusion that arsenic is released by reductive dissolution of iron oxyhydroxides, as was proposed earlier². Our observations indicate that arsenic-rich pyrite and other arsenic minerals, which were proposed in previous models (cited by Nickson *et al.*¹) to give rise to arsenic pollution, are rare or even absent in the sediments of the Ganges delta. We believe that arsenic is more likely to be co-precipitated with or scavenged by iron (III) and manganese (IV) in the sedimentary environment.

Nickson *et al.*¹ suggested that the arsenic in these alluvial sediments is derived from sulphide deposits in the Ganges basin. However, the copper belt of Bihar, which contains small amounts of arsenopyrite, and the coal basins of the Damodar valley, which contain moderate concentrations of arsenic, are drained by rivers that flow far to the south of the Ganges tributary system (Fig. 1). We suggest that there are several more likely sources of sedimentary arsenic, including the Gondwana coal seams in the Rajmahal basin, which contain up to 200 parts per million (p.p.m.) of arsenic; isolated outcrops of sulphides in the Darjeeling Himalayas, which contain up to 0.8% arsenic; and other sources in the upper reaches of the Ganges river system.

The Ganges alluvial tract upstream of Rajmahal, in the states of Bihar and Uttar Pradesh, does not suffer from large-scale arsenic contamination. This indicates that the Quaternary sediments around the Ganges delta have characteristic features that favour the initial retention and subsequent release of arsenic. These sediments have high proportions of clay and contain relatively large amounts of organic carbon, and are thicker towards the south^{3,4}. The average arsenic content in cores from boreholes in West Bengal is higher in layers of clay (9.5–12 p.p.m.) than of sand (3.8–4.8 p.p.m.)⁵. Quaternary sediments in the Ganges alluvial tract in Bihar and Uttar Pradesh contain more sand and are much narrower than sediments in the Bengal basin, which may explain why they retain less arsenic.

The groundwater of Uttar Pradesh and Bihar has trace concentrations of iron (0 to 0.7 mg per litre) compared with higher values in West Bengal (up to 36 mg per litre) and Bangladesh (30 mg per litre)¹.

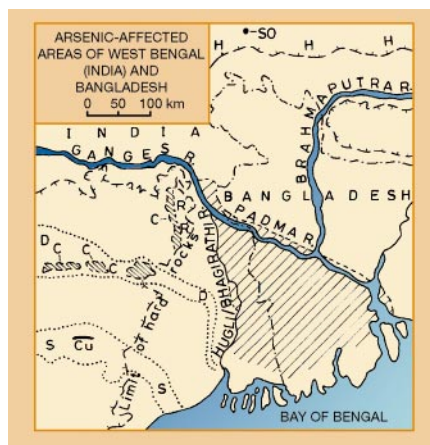


Figure 1 Map showing arsenic-affected areas (shaded). Cu, copper belt of Bihar; S and D, Subarnarekha and Damodar river basins; C, coal basins; R, Rajmahal volcanics; H, Himalayas; and SO, sulphide occurrence with arsenic in Darjeeling Himalayas.

The relatively low values of dissolved iron upstream of the Ganges delta indicate that the environment may not be sufficiently reducing to mobilize iron and arsenic.

Nickson *et al.*¹ reported that arsenic concentration increases with depth in wells at Manikganj, Faridpur and Tungipara in Bangladesh. However, this observation appears to be site specific, as the large database^{5–7} for aquifers in West Bengal indicates that arsenic decreases with depth (Fig. 2).

During the past thirty years, groundwater has been used increasingly for irrigation and the use of phosphate fertilizers has increased threefold. More than 0.5 million tubewells with handpumps, 0.1 million shallow tubewells, and 3,000 deep tubewells have been sunk at depths of 10–20 m, 30–100 m and 50–200 m, respectively^{8,9}. This widespread withdrawal of groundwater may have mobilized phosphate derived from fertilizers and from the decay of natural organic materials in shallow aquifers. The increase in phosphate concentration could promote the growth of sediment biota and the desorption of arsenic from sediments. These combined microbiological and chemical processes might have

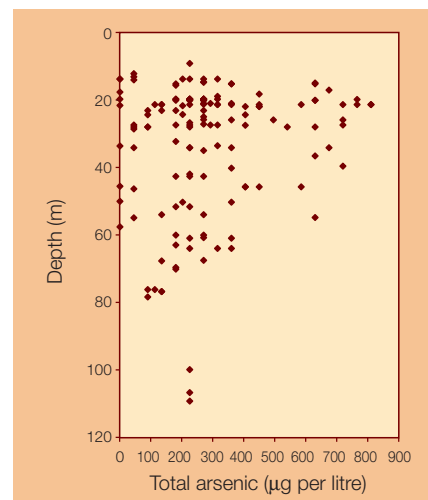


Figure 2 Variation of total arsenic with depth in aquifers in West Bengal.

increased the natural mobility of arsenic.

The efficiency with which arsenic is removed by adsorption onto iron-coated sand¹⁰ and by adsorption on and co-precipitation with ferrihydrite¹¹ depends on both the arsenic oxidation state and the ratio of iron to arsenic. The proposed removal of arsenic by simple aeration¹ of anoxic water must therefore be approached with caution.

S. K. Acharyya*, P. Chakraborty*, S. Lahiri*, B. C. Raymahashay†, Saumyen Guha†, Amitava Bhowmik†

*Geological Survey of India, Calcutta 700016, India

†Department of Civil Engineering, Indian Institute of Technology, Kanpur 208016, India

1. Nickson, R. *et al.* *Nature* **395**, 338 (1998).
2. Bhattacharya, P. *et al.* *Int. J. Water Res. Dev.* **13**, 79–92 (1997).
3. Hiller, K. *Geol. Jahrb. D* **90**, 3–35 (1988).
4. Mukherjee, A. & Hazra, S. *Ind. J. Geol.* **69**, 41–54 (1997).
5. Center for the Study of Man and the Environment, Calcutta (unpublished data).
6. Central Groundwater Board, New Delhi (unpublished data).
7. Geological Survey of India, Calcutta (unpublished data).
8. Bagla, P. & Kaiser, J. *Science* **274**, 174–175 (1996).
9. Mallick, S. & Rajagopal, N. R. *Curr. Sci.* **70**, 956–958 (1996).
10. Joshi, A. & Chaudhuri, M. *J. Env. Eng. ASCE* **122**, 769–771 (1996).
11. Wong, P. L. N., Huang, J. C. Y. & Cheng, T. W. J. *Chin. Inst. Env. Eng.* **5**, 241–251 (1995).

We have been studying the contamination of groundwater by arsenic and the attendant human suffering in West Bengal, India, for a decade, and in Bangladesh for the past four years. From our analysis of thousands of samples of water and sediment^{1–7}, we have been able to test the course of events proposed by Nickson *et al.*⁸ to account for the poisoning of

Bangladesh groundwater. We disagree with Nickson *et al.*'s claim that arsenic concentrations in shallow (oxic) wells are mostly below 50 µg per litre. In our samples from Bangladesh ($n=9,465$), 59% of the 7,800 samples taken at known depth and containing arsenic at over arsenic 50 µg per litre were collected from depths of less than 30 m, and 67% of the 167 samples with arsenic

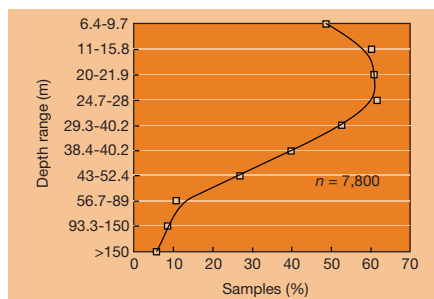


Figure 1 The proportion of well samples containing more than 50 µg per litre arsenic changes with depth in Bangladesh.

concentrations above 1,000 µg per litre were collected from wells between 11 and 15.8 m deep.

In the Lakshimpur district of Bangladesh, three of the shallowest tubewells (depths of 6.4 to 9.7 m) contained arsenic concentrations of over 1,000 µg per litre. The shallowest tubewell in Bangladesh is 6.4 m, and at that depth in the village of Chandipur, at the Ramganj police station, we found an arsenic concentration of 1,354 µg per litre. In the Noakhali district, we found arsenic at 2,700 µg per litre at a depth of 9.7 m. As shown in Fig. 1, many of the shallowest wells contain more than 50 µg per litre of arsenic. In West Bengal, ($n=55,000$), we found arsenic above 50 µg per litre in 28% of wells that were less than 34 m deep. Furthermore, 37 of 38 samples with more than 1,000 µg per litre of arsenic were collected from depths shallower than 34 m.

Nickson *et al.*⁸ reported that arsenic concentration increases with depth in wells at Manikganj, Faridpur and Tungipara, apparently on the basis of only a small number of samples⁹. From studies of 320 tubewells in these areas, we find that arsenic concentration increases with depth for depths of less than 22 m and decreases at depths of over 22 m (Fig. 1), and we have observed a similar trend in West Bengal. The *British Geological Survey* has recently reported that deep tubewells are free of arsenic in Bangladesh¹⁰.

We agree with Nickson *et al.* that arsenic associated with iron oxyhydroxides may be leached from the sediments under reducing conditions. However, we have suggested⁹ that this could be one of several ways in which arsenic is leached from aquifer sediments. We also proposed that the reducing environment in the younger region of the Ganges delta may have resulted in the formation there of arsenic-rich pyrite and that arsenic is mobilized by the oxidative dissolution of these solids^{3,6,7}.

This theory is based on our analysis of 2,235 sediment samples, taken at intervals of 3 or 6 m to a depth of 250 m, from 112 boreholes in West Bengal. Of these samples, 85 contained arsenic, ranging from 10 to 196 mg per kg. Opaque particles separated from the parent sediment contained arsenic^{3,6,7} at concentrations up to 2,778 mg

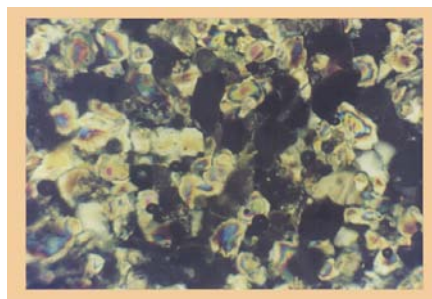


Figure 2 Photomicrograph (magnification, $\times 100$) showing the arsenic-containing opaque particles that are abundant in sediments. Analysis of selected opaque particles by X-ray diffraction and laser microprobe mass analysis identified the minerals pyrite, rozenite ($\text{FeSO}_4 \cdot 4\text{H}_2\text{O}$), haematite and magnetite in association with quartz and calcite.

per kg. Our microscopic examination of sediment samples from areas with arsenic-contaminated groundwater indicates an abundance of these opaque particles (Fig. 2). Electron microprobe analysis of the particles revealed increased iron, sulphur and arsenic^{3,6}, with arsenic content ranging from 0.07 to 1.36% by mass⁷.

Laser microprobe mass analysis identified arsenic-rich pyrite³, and X-ray diffraction identified the minerals pyrite, rozenite ($\text{FeSO}_4 \cdot 4\text{H}_2\text{O}$), haematite and magnetite in association with quartz and calcite^{3,7}. Pyrite was found⁷ in fine to coarse subangular/subrounded sands and occasionally with rock fragments at depths from 10 to 221 m. Arsenic has previously been found in pyrites from sediments of Bangladesh^{9,11}.

But the question remains over whether arsenic-rich pyrite is sufficient to account for the mass of arsenic mobilized in groundwater. It is often argued that if pyrite is oxidized, the aquifer should be rich in sulphate, but it is not. However, the formation of framboidal pyrite, which has been identified^{9,11} in Bangladesh, might explain the lack of sulphate in groundwater. Nonetheless, we are not convinced that a single mechanism is responsible for the release of arsenic from aquifer sediments in this region, and believe that both the pyrite-oxidation and iron oxyhydroxide-reduction hypotheses deserve further consideration.

We agree with Nickson *et al.* that groundwater in Bangladesh with high iron concentrations can be passively 'treated' by allowing oxidation, precipitation and settling of the ambient iron and the concomitant removal of arsenic. We have observed 60–70% arsenic removal in samples of well water in which a brown precipitate (presumably iron oxyhydroxide) had formed over time². However, the efficiency of arsenic removal depends on the oxidation of arsenic and the concentration of iron, and arsenic and iron occur together in some but not all samples. The proposed treatment would therefore not be effective for samples that are high in arsenic but low

in iron, as we and others¹⁰ have found in Bangladesh.

Tarit Roy Chowdhury*, Gautam Kumar Basu*, Badal Kumar Mandal*, Bhajan Kumar Biswas*, Gautam Samanta*, Uttam Kumar Chowdhury*, Chitta Ranjan Chanda*, Dilip Lodh*, Sagar Lal Roy*, Khitish Chandra Saha*, Sibtos Roy†, Saiful Kabir†, Qazi Quamruzzaman†, Dipankar Chakraborti*

*School of Environmental Studies,

Jadavpur University, Calcutta 700032, India

e-mail: dcoesju@vsnl.com

†Dhaka Community Hospital Trust,

Bara Moghbazar, Dhaka 1217, Bangladesh

1. Das, D. *et al.* *Analyst* **120**, 917–924 (1995).
2. Das, D. Thesis, Jadavpur Univ., Calcutta, India (1995).
3. Das, D. *et al.* *Environ. Geochem. Health* **18**, 5–15 (1996).
4. Dhar, R. K. *et al.* *Curr. Sci.* **73**, 48–59 (1997).
5. Samanta, G. *et al.* *Microchem. J.* **62**, 174–191 (1999).
6. Mandal, B. K. *et al.* *Sci. Tot. Environ.* **218**, 185–201 (1998).
7. Roy Chowdhury, T. Thesis, Jadavpur Univ., Calcutta, India (1999).
8. Nickson, R. *et al.* *Nature* **395**, 338 (1998).
9. Nickson, R. Thesis, University College London (1997).
10. Mott MacDonald Ltd *British Geological Survey Final Report* (January 1999).
11. Akai, J. *et al.* in *Conf. Proc. 3rd Forum Arsenic Contamination of Groundwater in Asia* 51–54 (Faculty of Engineering, Miazaki University, Japan, 1998).

McArthur replies — In their comments on our paper¹, both Acharyya *et al.* and Chowdhury *et al.* refer to previous suggestions^{2,3} that iron reduction might be the source of arsenic pollution. Naming a process², suggesting that it occurs on the basis of reasoned argument³, and providing objective evidence that it does^{1,4} are different things.

Unpublished data from both sets of authors confirm reports^{3–7} that concentrations of dissolved arsenic peak at depths of 20 to 40 m at some localities in the Ganges plain. These maxima appear in data aggregated from separate wells dispersed across unspecified areas, so their existence needs to be confirmed by vertical profiling at single localities. Nevertheless, these important data indicate that our findings that arsenic concentration increases with depth^{1,4} at some sites may not be a typical trend.

It has been reported⁶ that 1% of wells deeper than 200 m are contaminated with arsenic. We reported¹ that arsenic concentrations were mostly below 50 µg per litre in oxic (shallow) wells, not shallow (oxic) wells, as stated by Chowdhury *et al.*: the difference is that depth control on concentrations of dissolved arsenic is subordinate to redox control. The statement⁶ that "shallow hand-dug wells are usually uncontaminated" is not surprising as they are likely to be oxic. As neither Acharyya *et al.* nor Chowdhury *et al.* have provided data on the redox state of the water they analysed, their data throw no light on these findings.

Pyrite oxidation² is unlikely to be a major source of arsenic to anoxic water, which is the prevalent type beneath the southern Ganges plain. Oxidation of pyrite

by molecular oxygen yields acidity, whereas reduction of iron oxyhydroxide yields bicarbonate alkalinity. In anoxic water, arsenic correlates positively, not negatively, with bicarbonate¹, so it follows that the arsenic comes from the reduction of iron-oxyhydroxides. The presence of arsenical pyrite in aquifer sediments² shows that sulphate reduction occurred after, or in micro-environments at the same time as, iron reduction. The latter yield arsenic for incorporation into diagenetic pyrite⁸, which is a sink for, not a source of, arsenic.

Removing some of the arsenic from water by co-precipitation with iron oxyhydroxide (where possible) is a useful interim measure until better solutions^{9,10} are in place. Faced with a choice between using water that contains 250 µg per litre of poisonous arsenic or, after co-precipitation with iron, water that contains half that amount, who would not opt for the latter?

J. M. McArthur

Geological Sciences, University College London,
Gower Street, London WC1E 6BT, UK
e-mail: j.mcarthur@ucl.ac.uk

1. Nickson, R. *et al.* *Nature* **395**, 338 (1998).
2. Mandal, B. K. *Sci. Tot. Environ.* **218**, 185–201 (1998).
3. Bhattacharya, P., Chatterjee, D. & Jacks, G. *Water Res. Dev.* **13**, 79–92 (1997).
4. Nickson, R. *et al.* *Appl. Geochem.* (in the press).
5. Karim, M. *et al.* *J. Environ. Chem.* **7**, 783–792 (1997).
6. Mott MacDonald Ltd *British Geological Survey Final Report* (January 1999).
7. Maynard, D. M., Frisbie, S. H. & Hoque, B. A. NRECA Report (Bangladesh Rural Electrification Board, Dhaka, 1997).
8. McArthur, J. M. *Mineral. Mag.* **42**, 221–228 (1978).
9. Jekel, M. R. in *Arsenic in the Environment Part 1* (ed. Nriagu, J. O.) 119–132 (Wiley, New York, 1994).
10. Joshi, A. & Chaudhuri, M. *J. Environ. Eng.* **122**, 769–771 (1996).

Cell division

A histone-H3-like protein in *C. elegans*

The segregation of a chromosome during mitosis is mediated by a region of the chromosome known as the centromere, which organizes the kinetochore, to which the spindle microtubules attach. Many organisms have monocentric chromosomes, in which the centromeres map to single loci, whereas others, including the nematode *Caenorhabditis elegans*, have holocentric chromosomes, in which non-localized kinetochores extend along the length of each chromosome^{1,2}. The centromeres of monocentric chromosomes use specialized nucleosomes containing histone-H3-like proteins (known as CENP-A in mammals^{3–6} and Cse4 in the yeast *Saccharomyces cerevisiae*^{7,8}). Here we show that a *C. elegans* histone-H3-like protein is necessary for the proper segregation of chromosomes during mitosis and identifies the centromeres of these holocentric chromosomes, indicating that both holocentric and monocentric

chromosomes use centromeric histone-H3-like proteins.

The *C. elegans* protein encoded by F58A4.3 (ref. 9) is referred to here as HCP-3 (for holocentric protein-3). It is similar in sequence to histone H3 in its conserved histone-fold domain, but is dissimilar in its amino-terminal region, as are CENP-A and Cse4 (ref. 7). HCP-3 is the only *C. elegans* histone-H3-like protein represented by expressed sequence tags in the GenBank database. We therefore tested whether HCP-3 is a centromeric protein that is required for chromosome segregation.

To determine whether HCP-3 functions in mitosis, we used RNA-mediated interference¹⁰ to reduce *hcp-3* expression. This technique resulted in decreased staining with an anti-HCP-3 antibody: 87% ($n=46$ embryos) did not stain, compared with 4.9% ($n=61$ embryos) for non-injected worms. It also led to embryonic lethality in 100% of cases ($n=11$ adults) compared with 0.7% ($n=5$ adults) for non-injected worms, measured during a 24-hour period with adults that produced at least 20 embryos each. Staining with the DNA-binding dye DAPI revealed nuclei of varied size and staining intensity in the embryos with reduced HCP-3 (Fig. 1a,b), consistent with mis-segregation of chromosomes. Defects were evident as early as the first mitotic division. We conclude that HCP-3 is required for the proper segregation of chromosomes during mitosis.

Because CENP-A and Cse4 are only found at centromeres^{3,8}, we used anti-HCP-3 antibodies to visualize centromeres in *C.*

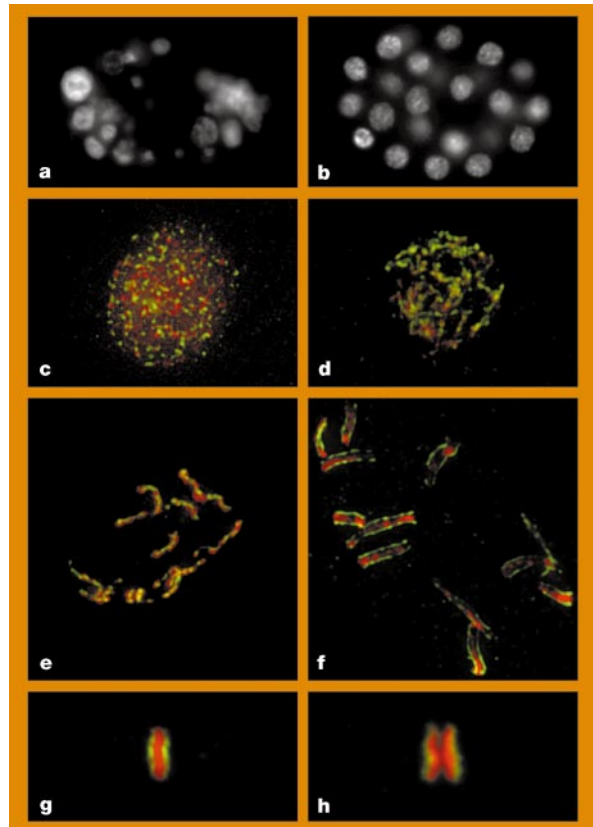
elegans. Anti-HCP-3 staining of interphase nuclei was punctate (Fig. 1c). In prophase, the antibody reacted with single lines associated with the condensing chromosomes (Fig. 1d,e). At prometaphase, there were two distinct, longitudinal bands of reactivity on each chromosome (Fig. 1f). We interpret this transition from one to two lines as a reorganization of the sister chromatids. At metaphase, the lines of reactivity were on the poleward faces of the metaphase plate (Fig. 1g). During anaphase, one set of single lines was associated with each set of separated sister chromatids (Fig. 1h).

The reorganization of HCP-3 between interphase and prophase indicates that dispersed regions of the chromosomes come together during chromosome condensation to form functional *C. elegans* centromeres. It has been suggested that the centromeres of mammalian chromosomes consist of discrete modules brought together in an orientation that aligns multiple microtubule-binding sites¹¹. The centromeres of holocentric chromosomes may make use of similar modules, using specialized centromeric nucleosomes, which are simply brought together from greater distances. Our results indicate that a broader analysis of holocentricity may lead to a better understanding of chromosome segregation.

Brian J. Buchwitz*, Kami Ahmad,
Landon L. Moore, Mark B. Roth*,
Steven Henikoff*†

Division of Basic Sciences, *Molecular and Cellular
Biology Program, and †Howard Hughes Medical
Institute, Fred Hutchinson Cancer Research Center,

Figure 1 A centromeric histone-H3-like protein from *Caenorhabditis elegans*. **a**, RNA-mediated interference of *hcp-3* resulted in mis-segregation of the chromosomes. **b**, A wild-type embryo, for comparison. For **a** and **b**, embryos were stained with the DNA-binding dye DAPI. **c–h**, The reorganization of HCP-3. **c**, Interphase; **d**, early prophase; **e**, prophase; **f**, prometaphase; **g**, metaphase (the holocentric chromosomes form a disc-like structure at metaphase that appears as a rod when viewed from the side); **h**, anaphase. For **c–h**, images were of nuclei from four-cell embryos, stained with DAPI (red) and anti-HCP-3 antibody (green). Scale bars, 10 µm. For **a**, **b**, **g** and **h**, embryos were fixed in *N,N*-dimethylformamide. For **c–f**, embryos were fixed in formaldehyde and a DeltaVision microscope with deconvolution software (Applied Precision, Issaquah, Washington) was used to generate projections. Details of methods are available from the authors.



1100 Fairview Avenue North, Seattle,
Washington 98109, USA
e-mail: mroth@fred.hcrcc.org

1. Albertson, D. G., Rose, A. M. & Villeneuve, A. M. in *C. elegans*, Vol. II (eds Riddle, D. L. et al.) 47–78 (Cold Spring Harbor Laboratory Press, Plainview, New York, 1997).
2. Albertson, D. G. & Thomson, J. N. *Chromosoma* **86**, 409–428 (1982).
3. Earnshaw, W. C. & Rothfield, N. *Chromosoma* **91**, 313–321 (1985).
4. Palmer, D. K. et al. *J. Cell Biol.* **104**, 805–815 (1987).
5. Palmer, D. K. et al. *Proc. Natl Acad. Sci. USA* **88**, 3734–3738 (1991).
6. Sullivan, K. F., Hechenberger, M. & Masri, K. *J. Cell Biol.* **127**, 581–592 (1994).
7. Stoler, S. et al. *Genes Dev.* **9**, 573–586 (1995).
8. Meluh, P. B. et al. *Cell* **94**, 607–613 (1998).
9. Wilson, R. et al. *Nature* **368**, 32–38 (1994).
10. Fire, A. et al. *Nature* **391**, 806–811 (1998).
11. Zinkowski, R. P., Meyne, J. & Brinkley, B. R. *J. Cell Biol.* **113**, 1091–1110 (1991).

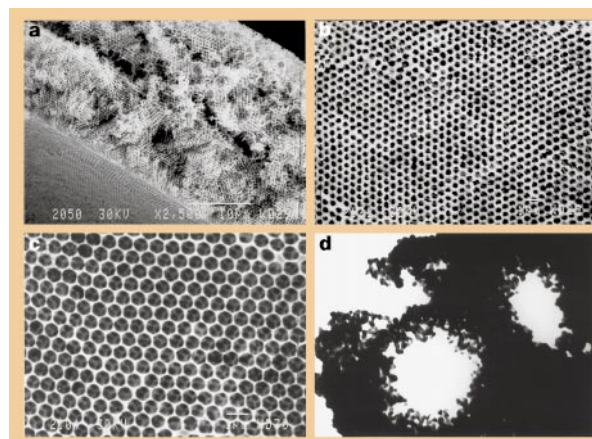
Materials

A class of porous metallic nanostructures

Colloidal crystals are ordered arrays of particles in the nanometre-to-micrometre size range. Useful microstructured materials can be created by replicating colloidal crystals in a durable matrix that preserves their key feature of long-range periodic structure¹. For example, colloidal crystals have been used to fabricate structures from inorganic oxides^{1–5}, polymers^{6,7}, diamond and glassy carbon⁸, and semiconductor quantum dots⁹, and some structures have photonic properties^{4,8,9} or are patterned on different hierarchical length scales⁵. By using colloidal crystals as templates, we have synthesized a new class of metallic materials with long-range nano-scale ordering and hierarchical porosity.

We formed porous metallic nanostructures by using colloidal crystal templates from monodisperse, negatively charged polystyrene latex microspheres 300 nm to 1 µm across. The close-packed crystals were assembled by filtration^{1,3} through a smooth polycarbonate membrane with pores (50 nm) that retained the latex and the gold particles but allowed a high flux of water. The diluted latex particles were slowly deposited into densely packed layers about 35 µm thick. Colloidal gold particles 15–25 nm across were then slowly deposited in the

Figure 2 Electron micrographs of the structure of porous gold obtained with 630-nm latex templates. **a**, Side view of the flake, showing long-ranged ordered arrays of pores, both on the surface and through the flake. **b**, A typical area on the surface of a calcined sample, showing fusion, domain boundaries and crystal defects. **c**, A sample obtained by latex dissolution. **d**, Transmission electron micrograph at the edge of a flake obtained by dissolution, showing the mesoporous particulate gold structure.



interstices of the latex crystals by filtration, forming a mesoporous structure around the latex microspheres. This porosity allowed the solvent to flow through the deposit until the pores were filled by the gold colloid. Because of diffraction, the dried latex–gold flakes generate strongly coloured reflections that are much brighter than those of the original latex templates alone (Fig. 1a, b).

The three different procedures we used to remove the latex beads from within the composite yielded porous metal with different properties. Calcination for 30 min at 300 °C yielded flakes with the colour of metallic gold, but with multicoloured domains on the surface (Fig. 1c). Removing the latex templates at room temperature by oxidation with acid or by dissolving in trichloromethane gave brown flakes with brightly coloured reflections. All the samples were conductive, with resistance higher than, but of the same order of magnitude as, that of metallic gold foil.

Scanning and transmission electron microscopy revealed that the materials have a three-dimensional structure of ordered pores throughout the flakes (Fig. 2a). As in previous observations¹, monocrystalline hexagonal domains of up to a thousand identical pores are predominant on the surface, usually corresponding to three-dimensional, randomly stacked, hexagonal close-packed planes and face-centred cubic packing (Fig. 2b,c), although square arrays are occasionally seen. The lattice dimensions correspond to those of the original latex crystals, indicating that there was negligible

shrinkage of the metallic structure, in contrast to that seen for inorganic oxides^{1–4}.

The materials formed by calcination are macroporous, with a smooth, fused gold structure (Fig. 2b). In contrast, the solvent dissolution method preserves the original particulate gold structure around the large pores (Fig. 2d). In this mesoporous–macroporous material, the size of the small pores and the overall surface area are determined by the size of the gold particles. Materials with such controlled hierarchical porosity may be particularly suitable for catalytic applications, as they provide a combination of efficient transport and high surface area.

The lace-like structure of the flakes (Fig. 2c) is an almost exact replica of some three-dimensional wire-mesh photonic crystals¹⁰, but reduced by a factor of 20,000 to the sub-micrometre scale. The materials may therefore have photonic properties in reflectance mode in the visible-light region. In principle, the method can be improved to allow the formation of layers only a few pores thick, with interesting transmission properties. The simple ‘wet’ method can also be adapted to create materials from other metals, alternating layers of metals, or composite metallic–dielectric structures. Such new materials could be used in advanced catalysis or electro-optical devices, as well as in reflective, conductive or energy-collecting metallic coatings.

O. D. Velev, P. M. Tessier, A. M. Lenhoff, E. W. Kaler

Center for Molecular and Engineering
Thermodynamics, Department of Chemical
Engineering, University of Delaware,
Newark, Delaware 19716, USA
e-mail: velev@che.udel.edu

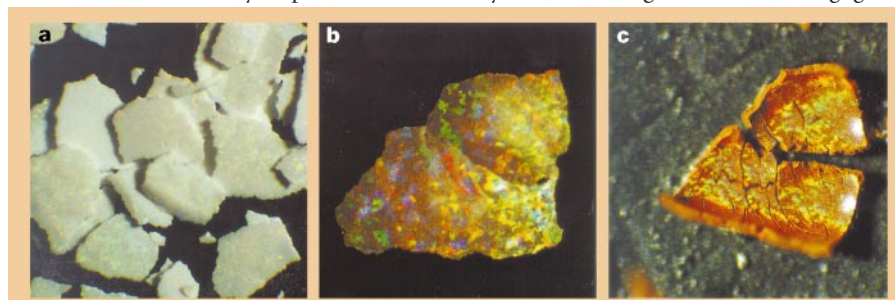


Figure 1 Optical micrographs of the materials formed with 630-nm latex at different stages of synthesis, in reflected polychromatic illumination. **a**, Original latex colloidal crystal templates. **b**, Latex–gold composite before removal of the template. **c**, Gold-only flake after calcination. Field of view: **a**, **b**, 1,800 µm; **c**, 900 µm. Further details of the methods are available from the authors.

1. Velev, O. D. et al. *Nature* **389**, 447–448 (1997).
2. Imhof, A. & Pine, D. J. *Nature* **389**, 948–951 (1997).
3. Holland, B. T., Blanford, C. F. & Stein, A. *Science* **281**, 538–540 (1998).
4. Wijnhoven, J. E. G. J. & Vos, W. L. *Science* **281**, 802–804 (1998).
5. Yang, P. et al. *Science* **282**, 2244–2246 (1998).
6. Park, S. H. & Xia, Y. N. *Chem. Mater.* **10**, 1745–1747 (1998).
7. Johnson, S. A., Ollivier, P. J. & Mallouk, T. E. *Science* **283**, 963–965 (1999).
8. Zakhidov, A. A. et al. *Science* **282**, 897–901 (1998).
9. Vlasov, Y. A., Yao, N. & Norris, D. J. *Adv. Mater.* **11**, 165–169 (1999).
10. Sievenpiper, D. F., Sickmiller, M. E. & Yablonovitch, E. *Phys. Rev. Lett.* **76**, 2480–2483 (1996).

Multi-gas assessment of the Kyoto Protocol

J. Reilly*, R. Prinn*, J. Harnisch*, J. Fitzmaurice*, H. Jacoby*, D. Kicklighter†, J. Melillo†, P. Stone*, A. Sokolov* & C. Wang*

* Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building E40, Cambridge, Massachusetts 02139, USA

† The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

The Kyoto Protocol allows reductions in emissions of several 'greenhouse' gases to be credited against a CO₂-equivalent emissions limit, calculated using 'global warming potential' indices for each gas. Using an integrated global-systems model, it is shown that a multi-gas control strategy could greatly reduce the costs of fulfilling the Kyoto Protocol compared with a CO₂-only strategy. Extending the Kyoto Protocol to 2100 without more severe emissions reductions shows little difference between the two strategies in climate and ecosystem effects. Under a more stringent emissions policy, the use of global warming potentials as applied in the Kyoto Protocol leads to considerably more mitigation of climate change for multi-gas strategies than for the—supposedly equivalent—CO₂-only control, thus emphasizing the limits of global warming potentials as a tool for political decisions.

Many trace atmospheric constituents affect the radiative budget of the atmosphere. The Kyoto Protocol includes carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), perfluorocarbons (PFCs), hydrofluorocarbons (HFCs), and sulphur hexafluoride (SF₆) (<http://www.unfccc.de/resource/docs/convkp/kpeng.pdf>). Some studies suggest that potential carbon sinks^{1,2} and opportunities to limit emissions of other greenhouse gases^{3–5} may reduce the cost of control. These also point to the risks of failing to control gases with very long lifetimes^{6,7}. Few studies have yet considered an integrated evaluation of the costs of multi-gas control strategies, or the implications of reductions in different mixes of greenhouse gases on atmospheric composition, climate and ecosystems^{8–11}.

Using the MIT Integrated Global System Model¹² (IGSM) as recently modified¹³ we examine multi-gas control as envisioned by the Kyoto Protocol, exploring the costs of emissions reduction and the consequences for the atmosphere, climate and ecosystems. The basic components of the IGSM are an emissions prediction and policy analysis (EPPA) model¹⁴, a natural emissions model, a coupled atmospheric chemistry and climate model^{15,16}, and a terrestrial ecosystems model¹⁷. We stop short of quantifying damages in monetary terms. Our ecosystems results illustrate trade-offs that result from different control strategies. We find that inclusion of sinks and abatement opportunities for gases other than CO₂ could reduce the cost of meeting the Kyoto agreement by 60%. Assuming the protocol is extended unchanged to 2100, we find little difference in climate and ecosystem effects between 2010 and 2100 for a strategy that achieves the required reduction with a multi-gas strategy as compared to a CO₂-only strategy. Under a more aggressive policy, increasing the reductions in Annex B countries (Eastern and Western Europe, Russian Federation, Australia, New Zealand, Japan, Canada and the United States) and extending reductions to the rest of the world after 2010, significant differences in effects develop between the two strategies. This latter result indicates that 100-year global warming potentials (GWPs), as currently estimated, fail to capture important time horizon and climate–chemistry interactive effects, and this failure may be significant for policy.

Policy cases

The Kyoto Protocol establishes maximum allowable emissions of greenhouse gases aggregated using 100-year GWPs for the period 2008 to 2012 for each party in Annex B of the agreement, stated as a percentage of base year emissions. Global warming potentials are

indices that supposedly establish equivalency in terms of climate effects for purposes of comparing reductions in emissions of different greenhouse gases to assess compliance with the Kyoto agreement¹⁸. The motivation is that such gases have very different lifetimes and specific radiative forcing and thus are not at all comparable on a tonne-for-tonne basis. The Kyoto Protocol also allows credit for carbon sinks resulting from direct, human-induced afforestation and reforestation measures occurring after 1990. The Protocol rules out credit for indirect sink enhancement from increased growth of plants due to increased atmospheric CO₂ itself or from deposition of nitrogen emitted from industrial processes. Thus, although the role of terrestrial ecosystems in balancing the carbon budget is important in understanding the carbon cycle¹⁹ the magnitude of a natural land sink is irrelevant to Kyoto accounting. Inclusion of agricultural soils and a broader set of afforestation activities are under discussion.

Global warming potentials hide complex feedbacks among ecosystems, the atmosphere, and the oceans, and omit responses that are nonlinear with changing levels and rates of increase of different gases. They also ignore chemical interactions among the greenhouse gases and other gases, including processes of formation and destruction that depend on climate itself¹⁵. Patterns of forecast climate change may differ under alternative combinations of gas controls because spatial and transient forcing responses can be heterogeneous²⁰. Also, GWPs depend on the choice of time horizon for integration of the effects of radiative forcing. For example, defined relative to CO₂ (CO₂ = 1.0), the shorter-lived CH₄ has a higher GWP with a shorter time horizon (21 under a 100-year horizon, 56 under a 20-year horizon)¹⁸. An alternative to using a finite time horizon is to formulate GWPs as a solution to an infinite horizon economic problem with a discount rate to provide an economic weighting of gases over time. Global warming potentials would then depend on the actual effects of climate change and greenhouse-gas accumulation on the economy, human wellbeing, and ecosystems. Such a calculation requires effects to be estimated in a comparable measure, usually in monetary (for example, US dollar) terms^{21–23}. Carbon dioxide also has direct effects on plant growth that should be included in damage estimates. Their inclusion would further affect GWP calculations²². There are significant limitations on existing efforts to value damages in monetary terms²⁴, however, and controversy about the use of discounting in climate change policy analysis²⁵.

In addition to a future reference case with no emission controls,

we develop three basic policy cases to test the economic importance of including non-CO₂ gases in the Kyoto Protocol.

Case 1. Fossil CO₂ target and control. This case is representative of much previous work on the costs of limiting the greenhouse effect. It includes only CO₂ in determining the allowable emissions, unlike the requirements in the Kyoto Protocol that consider multiple gases.

Case 2. Multi-gas target with control on CO₂ emissions only. This case is constructed with the multi-gas target (expressed as carbon equivalents using GWPs), as described in the Kyoto Protocol, but only carbon emissions from fossil fuels are controlled.

Case 3. Multi-gas target and controls. The multi-gas Kyoto target applies in this case, and parties seek the least cost control across all gases and carbon sinks.

For purposes of analysis of climate and ecosystem effects over the longer term we extend policy cases 2 and 3 under two sets of assumptions. In one we assume that the Kyoto commitment remains unchanged to the end of 2100 (extended cases 2 and 3). We also define much more stringent versions of cases 2 and 3, denoted cases 2' and 3' respectively. In these latter cases Annex B reduces carbon-equivalent emissions by an additional 5% in each succeeding 15-year period, to 35% below 1990 levels by 2100. We assume that non-Annex B regions join in 2025, adopting a target of 5% below their 2010 emissions. They then follow Annex B by tightening the constraint by 5% every 15 years, reaching 30% below 2010 emissions in 2100.

We assume in case 3 that non-CO₂ gas abatement beyond 2010 is the same percentage below base levels as calculated for 2010, based on our economic assessment for 2010 and assuming also that the contribution from sinks remains unchanged through 2100. Case 2 maintains the same carbon-equivalent emissions with CO₂-only controls, as calculated using 100-year GWPs. For case 3' we assume that reductions beyond 2010 for all gases are proportional to the reduction commitment, and case 2' again keeps the same carbon-equivalent path, with control on CO₂ only. The purpose of the extended cases is to examine whether the choice of gases to control makes a difference to the climate system. Cases 2' and 3' in particular differ significantly in emissions of both CO₂ and other gases. As a result, they provide a good test of the climate and ecosystem implications of using the 100-year GWPs with different strategies for implementing the agreement.

Figure 1a shows that the difference between anthropogenic CO₂ (plus CO, which converts to CO₂) emissions and total natural CO₂ sinks (predicted for the ocean and prescribed for the land¹²) grows with time in the reference case and cases 2 and 3 but stabilizes or falls in cases 2' and 3'. Reference emissions of carbon reach 17.4 gigatonnes (Gt) by 2100, lower than the previous EPPA reference of 20.6 Gt (ref. 12). When combined with other greenhouse gas emissions using 100-year GWPs, total emissions reach about 25 Gt C equiv. (gigatonnes of carbon equivalent) by 2100 (Fig. 1b). Under cases 2 and 3, combined annual emissions are reduced by about 7 Gt C equiv. in 2100, whereas emissions are reduced by about 17 Gt C equiv. in cases 2' and 3'.

Economic assessment

Our economic analysis is limited to provisions of the Kyoto Protocol directed at Annex B countries, and focuses on the year 2010 as indicative of the annual cost in the Kyoto commitment period of 2008–12. Our approach uses marginal abatement curves (MACs) that allow the options available for abating emissions to be ordered from lowest cost to highest cost. The significance of this ordering is that an economically efficient policy would choose the lowest-cost option first, moving to higher-cost options as more abatement was required. Figure 2 shows MACs for CO₂ and all greenhouse gases for the USA, illustrating our approach and how it relates to the policy cases we examine. Marginal abatement curves for individual gases and sinks are shown in the inset. For the USA, the required reduction under case 1 (RR1) of 571 Mt C equiv.

(megatonnes of carbon equivalent) results in a price (P1) of \$187 per tonne of carbon equivalent (t C equiv.). With a multi-gas target the required reduction (RR2,3) rises to 650 Mt C equiv. If this target must be achieved with controls on only CO₂ emissions, then the price rises to \$229 per t C equiv., but taking advantage of abatement options for all gases and sinks reduces the price to \$150 per t C equiv.

The base year (1990) emissions, 2010 reference, and allowable emissions under the policy cases representing the Kyoto Protocol in our analysis (in carbon equivalent using 100-year GWPs) are presented in Table 1, along with the required reductions. Reference emissions are projected in our analysis to increase substantially by 2010 in all regions except the former Soviet Union (FSU). The increases from the base year range from 29% in the European Union (EEC) to 46% in Eastern Europe (EET). Aggregate emissions in the FSU are projected to be 7.5% less in 2010 than in 1990, most of this reduction having already occurred before 1998. The allowable emissions for the FSU exceed our reference emissions in 2010, giving the FSU 'hot air' that can be traded or banked against future increases in emissions. Including the non-carbon gases in calculating our reference emissions reduces this 'hot air' by nearly 60%, reflecting the fact that non-CO₂ gases in the FSU are projected to grow in our reference scenario. This reduction is mainly due to accounting for methane emissions. However, our model includes no direct measurements of methane emissions or abatement potential in the FSU region. 'Hot air' calculations are, in general, highly

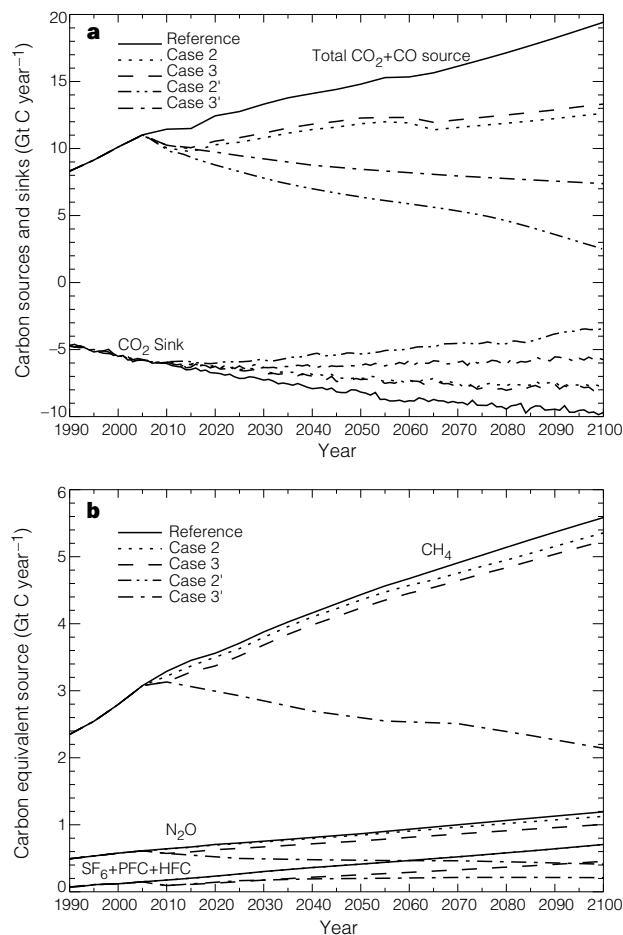


Figure 1 Future greenhouse gases emissions. Panel **a** shows total CO₂ and CO (which converts to CO₂) sources (positive) and total CO₂ sinks (negative) as references, and cases 2, 2', 3 and 3'. Panel **b** shows non-CO₂ sources expressed as equivalent amounts of CO₂ emissions using global warming potentials with 100-year horizons. The case 2' plot lies on top of the reference plot in panel **b**.

uncertain. Another recent forecast of carbon emissions for the FSU implies that 'hot air' in 2010 will range up to 388 Mt (ref. 26).

We computed MACs for CO₂ reduction in 2010 using the EPPA model. Those for sinks and gases other than CO₂ were developed from detailed evaluations of economic reduction potential drawn from published literature, recent studies, and cost estimates of industry experts. Parameter estimates for the curves are given in Fig. 1. Marginal abatement curves for sinks were developed based on a US study of forest carbon sequestration on agricultural land¹. Quantities of sequestration for different regions available in 2010 were based on tree growth potential on currently idle or underused land, with regional MACs shifted horizontally to fit these regional quantities²⁷. We phased in the tree-planting programme as if it started in the year 2000.

The potential for CH₄ reduction was based on an assessment of the costs of recovery from landfills, livestock waste, coal seams, and oil and gas production conducted for the USA^{28,29}. Other EPPA regions were assumed to face similar costs for comparable shares of abatement of CH₄ emissions from the reference level. No CH₄ abatement is included for ruminant animals or rice production. Existing economic studies show the marginal costs to be very high for these latter sources¹, although these studies do not consider technological options that have been suggested more recently³⁰. The reduction potential for N₂O from fertilized soils was based on econometric studies of the price response of fertilizer demand^{31,32}. Additional reduction potential is assumed to arise from switching

between different types of synthetic fertilizers. Abatement from adipic and nitric acid production was based on industry estimates.

Estimates of the reduction potential and cost for other gases were based on information from the published literature and industry experts. Two main sources of HFCs are CHF₃ as a by-product of hydrochlorofluorocarbon (HCFC) manufacture³³ and CH₂FCF₃ from mobile air conditioning³⁴. Their use in stationary cooling, foam blowing, and solvents is likely to grow rapidly. We considered recovery and thermal oxidation of by-products of HCFC production and replacement of HFCs with CO₂ in mobile air conditioners (J. Wertenbach and R. Caesar, personal communication), and recovery and replacement options for the remaining sources. Perfluorocarbons such as CF₄ are emitted during production of primary aluminium⁵ and semiconductors, and through their use as a replacement for ozone-depleting substances phased out under the Montreal Protocol. We estimated reduction rates and costs for the aluminium industry⁵. Also, costs of recovery from gas streams and subsequent combustion, and costs of substitute chemicals that were assumed to be available to the semiconductor sector and to other miscellaneous sectors were estimated. Expert estimates were the basis for the costs of removal of PFCs from waste gases in the semiconductor industry and for the costs of substitute chemicals in solvent applications.

Sulphur hexafluoride is emitted from magnesium and semiconductor production, the manufacture and use of electrical switchgear⁴, and a broad category of other applications³⁵. For losses in switchgear manufacturing we account for the fact that recycling of SF₆ is economical at current prices, and assume rates of post-installation control at higher prices. Similarly, it is assumed that, at increased prices, magnesium producers will be able to cut specific emissions to values already achieved by Norwegian manufacturers³⁵.

The MACs for 2010 then allow study of the policy cases presented in the previous section. Case 1 allows comparison with previous studies that have considered only CO₂. Compared with case 3 it shows an overestimate of the carbon price ranging from about 8% in the EET (\$11) to 153% in the remainder of the OECD (OOE) (\$158); costs throughout are in 1985 \$US. Much of the difference comes from variation in the carbon sink potential among the regions. For example, there is relatively small sink potential in the EET and Japan and a very large sink potential relative to carbon emissions in the OOE. In terms of total cost, we find that, for Annex B as a whole, the error from leaving out other greenhouse gases is about a \$27 billion per year overestimate of the annual cost in 2010 (Table 2). The OOE, USA and EEC have substantially lower costs under case 3 than case 1. Notably, however, case 3 achieves a greater reduction in greenhouse gases (1.55 versus 1.34 Gt C equiv. per year) than case 1.

Cases 2 and 3 were designed to be comparable in terms of carbon-equivalent emissions abated; differences in costs thus reflect different options for implementing the agreement. We find that the carbon-equivalent price is over \$100 per tonne higher for the USA, EEC, OOE and Japan in case 2 than in case 3. In terms of total annual cost the difference between cases 2 and 3 is striking. To achieve the 1.55 Gt C equiv. reduction only through fossil energy carbon emissions would cost nearly \$62 billion per year more than using a multi-gas strategy with sink enhancement, increasing the cost of the agreement by over 60%. The EET is the only region with higher costs when other greenhouse gases and sinks are included. This happens because of their limited potential for sinks and the fact that emissions for non-CO₂ greenhouse gases are growing rapidly in the reference case. The multi-gas strategy in the Kyoto Protocol thus creates an opportunity for lower abatement costs, but introduces a risk of higher costs if reductions in these other gases are not made.

Some studies show a potential even larger than we have estimated for reducing non-CO₂ greenhouse gases³⁶. Leakage rates of CH₄ from gas and coal mine operations in countries of the FSU are

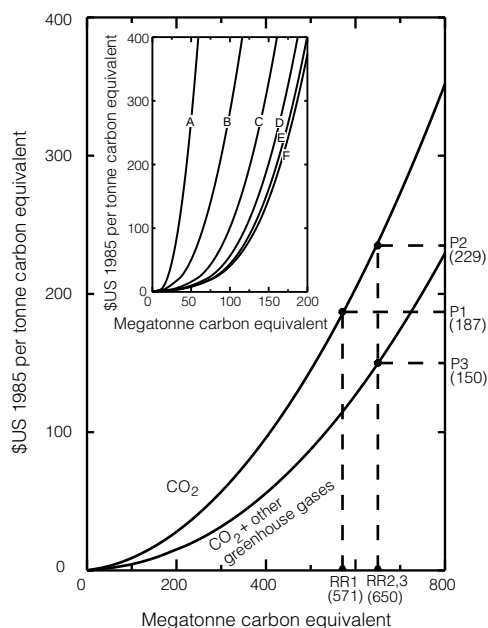


Figure 2 Marginal abatement curves (MACs). Calculated for CO₂ in the USA and total greenhouse gases. RR1, RR2,3 and P1, P2, P3 are the required reductions and prices under policy cases 1, 2 and 3. Inset provides sequentially summed MACs for other greenhouse gases in the USA in carbon equivalent units using 100-year global warming potentials. A: CH₄; B: CH₄ + carbon sink; C: CH₄ + sink + N₂O; D: CH₄ + sink + N₂O + HFCs; E: CH₄ + sink + N₂O + HFCs + SF₆; F: CH₄ + sink + N₂O + HFCs + SF₆ + PFCs. Marginal abatement curves for the forest sink and CH₄ are quadratic polynomials ($P = a + bQ + cQ^2$) where Q is the reduction in Mt C equiv. and P is the price in 1985 US dollars. For sinks: $a = 33.5$, $b = -3.59$, $c = 0.177$; for CH₄: $a = 12.15$, $b = -2.42$, $c = 0.15$. Parameters for other regions are available from the authors upon request. Other MACs are exponential functions ($P = a \exp(bS)$) where S is the share of reference emissions reduced. For N₂O: $a = 0.580$, $b = 0.177$; for PFCs: $a = 6.932$, $b = 0.036$; for HFCs: $a = 1.189$, $b = 0.060$; for SF₆: $a = 0.795$, $b = 0.065$, identical across regions. P3 given here is without "free" reductions of CH₄ and N₂O generated by the carbon constraint.

Table 1 Emissions scenarios

Gas	USA	EEC	OOE	EET	JPN	FSU
Base year (1990) emissions; (2010 Reference forecast)*						
CO ₂	1,362 (1,838)	822 (1,064)	318 (472)	266 (395)	298 (424)	891 (763)
CH ₄	170 (184)	129 (143)	69.9 (84.8)	64.5 (82.1)	9.0 (10.9)	155 (209)
N ₂ O	92.3 (121)	71.7 (92.3)	23.4 (31.9)	8.9 (10.6)	9.3 (13.7)	11.3 (14.2)
SF ₆	10.5 (12.7)	6.2 (7.7)	3.3 (4.2)	1.0 (1.6)	3.5 (4.9)	3.2 (3.5)
HFC	12.3 (27.6)	7.6 (17.7)	2.5 (4.2)	0.5 (1.7)	4.6 (11.6)	1.6 (4.6)
PFC	7.4 (5.0)	4.1 (2.8)	7.2 (5.1)	1.1 (0.9)	1.3 (0.3)	6.2 (5.3)
Total	1,654 (2,188)	1,042 (1,327)	425 (604)	342 (492)	326 (466)	1,068 (1,000)
Kyoto reduction (below 1990) %	-7.0	-8.0	-5.5	-7.0	-6.0	-0.2
Allowed anthropogenic emissions; (required reductions)*						
Case 1	1,267 (571)	757 (308)	301 (171)	248 (147)	280 (144)	889 (-126)
Cases 2 and 3	1,539 (650)	958 (369)	401 (203)	318 (174)	306 (160)	1,066 (-66)

* in Mt C equiv. (metatonnes of carbon equivalent).

Regions listed are those with countries that are included in Annex B of the Kyoto Protocol; the United States (USA), 12 countries of the European Union as of 1990 (EEC), Japan (JPN), the remainder of the OECD (OOE), the regions of the former Soviet Union (FSU) and Central and Eastern Europe (EET). Some of the FSU Republics are not part of Annex B. The base year for SF₆, PFCs, and HFCs is 1995. Some countries of the EET have chosen to use base years other than 1990. If recalculated with these alternative base years allowed emissions (required reductions) are 274 (121) Mt C equiv. in case 1 and 342 (150) Mt C equiv. in cases 2 and 3.

reported to be high, and it may well be economic to reduce these leakage rates even without climate policy. If so, these reductions in the FSU could be less expensive than represented in our analysis. The cost savings could thus be even larger and the scope for trade in emissions permits would be expanded. 'Hot air' would also then increase substantially beyond our estimate. Further refinement of region-specific costs is needed in order to understand the implications for trading opportunities better. Although the above caveats suggest lower marginal abatement costs, realization of the potential for reductions in other greenhouse gases and increased sinks requires that policies and economic incentives be put in place to bring these reductions about.

Our analysis assumes policies that achieve reductions at least cost. We have included 'no-cost' reduction options in our reference

scenario. For other greenhouse gases, our marginal abatement curves are based on 'bottom-up' engineering studies that sometimes fail to include some economic costs and barriers to implementing reductions. In addition, measuring and monitoring sinks and emissions from very diverse sources of greenhouse-gas reductions could add to cost, or require a larger economic incentive than we have estimated, for emitters to undertake abatement measures. This set of concerns, many of which fall into the category of transaction costs, could mean that our marginal abatement curves underestimate costs.

Carbon dioxide emissions are also uncertain and depend on a number of assumptions and sensitivities, as has been shown elsewhere^{12,37}. The handling of technical change and costing of alternative technologies has generated widely varying estimates of

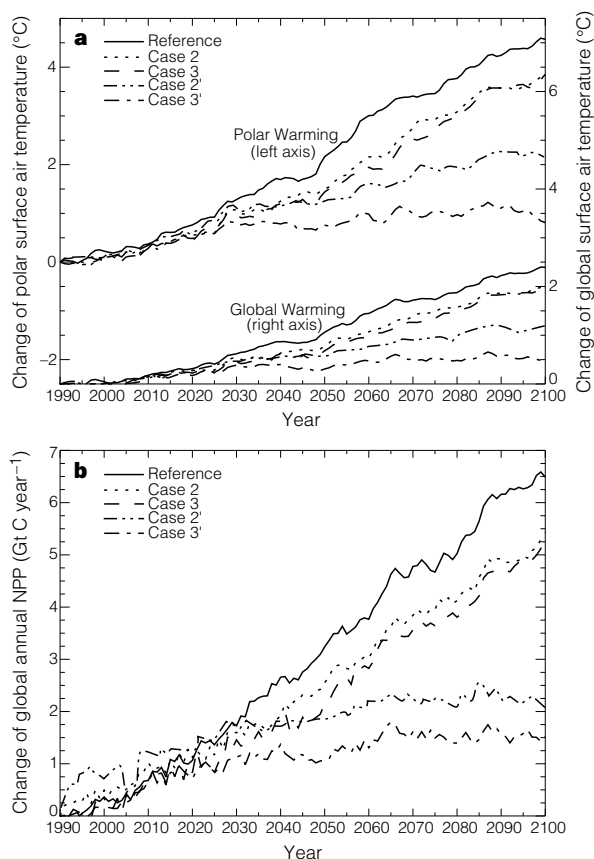


Figure 3 Climate and ecosystem effects. Panel **a** shows the three-year moving-average global and polar temperature changes from 1990 levels. With the Kyoto Protocol extended to 2100 (cases 2 and 3) there is a 17% reduction in warming from the reference. With more aggressive policies after 2010 (cases 2' and 3') warming is reduced by about 55 and 85% respectively from the reference. Panel **b** shows the three-year moving average of net primary production changes from 1990 levels. With the Kyoto Protocol extended to 2100 (cases 2 and 3), the net primary production increases by 22% less than in the reference. With a more stringent policy (cases 2' and 3') the increase is about 65 and 77% less respectively.

Table 2 Total annual abatement cost in 2010 (billions of 1985 \$US)

	USA	EEC	OOE	EET	JPN	FSU
Case 1: CO ₂ target and control	37.5	30.0	15.6	8.9	34.0	0
Case 2: Multi-gas target, CO ₂ control	45.6	44.2	18.2	11.4	42.2	0
Case 3: Multi-gas target and control	27.8	23.7	8.0	9.2	30.8	0

Regional definitions are given in Table 1. Costs are measured as the integral from no reduction to the projected reduction level under our marginal abatement curves. Notionally, costs are the foregone economic output associated with the greenhouse gases policy, as the carbon constraint is placed on producers and household consumers in the economic model. This cost measure is consistent with measurement of costs in the exogenous marginal abatement curves. Costs for the EET, using alternative base years (see Table 1) are 5.0, 7.9 and 6.0 billion dollars in cases 1, 2 and 3, respectively.

costs, some suggesting little or no cost for meeting the Kyoto target³⁸. These costing approaches are not always complete or consistent with a fuller treatment of economic costs³⁹. Recently attention has focused on the potential for a carbon constraint to induce technical change⁴⁰, showing potentially large effects depending on the representation of the process. These differences can lead to cost differences or $\pm 50\%$ in 2010. Although the base cost estimate is uncertain, and marginal abatement curves themselves are uncertain, we would expect the basic conclusion, that there is substantial potential to reduce costs through inclusion of other gases, to hold. The comparison and ordering of case 1 versus case 3 is less certain and is more dependent on the varying growth of emissions of each gas and the sink potential. For example, delay beyond 2000 in starting a forestation programme could mean that very little sequestration is achieved by 2010.

Climate effects

The new IGSM chemistry–radiation submodel used here¹³ yields a reference run with a predicted temperature increase from 1990 to 2100 of 2.4 °C. This is similar to the 2.6 °C in previous reference runs¹² despite several changes in EPPA and chemistry–radiation submodels. This is because the effects of the lower CO₂ emissions predicted by EPPA are offset approximately by the emissions of HFCs, PFCs, and SF₆ that were not included previously, whereas effects of the inclusion of the (predicted) increases in tropospheric ozone are offset approximately by higher SO_x emissions.

Predicted warming (and predicted sea-level rise) between 1990 and 2100 is lowered from the reference case by only $\sim 17\%$ in cases 2 and 3 (Fig. 3a) and there is little difference between the cases. This small difference indicates that, for the Kyoto policy extended through 2100, the crude approximation represented by the GWPs does not give rise to substantial differences in temperature effects. The little difference that exists is difficult to distinguish from the natural variability simulated by the climate model; it illustrates the effects, in cases 2 and 3, of interactions (chemical and radiative) between the non-CO₂ gases not captured by GWPs (such as the effects of added CO and CH₄ on their principal sink (OH radicals) and hence on the lifetimes of these two gases⁴¹). The absolute and percentage reduction in predicted global average warming is modest in cases 2 and 3, but the effect on absolute warming between 1990 and 2100 in the regions poleward of 51°N and 51°S is more substantial (Fig. 3a). Specifically, the average polar warming is lowered from about 4.6 °C in the reference run to about 3.8 °C in the case 2 and 3 runs, which may be important for the stability of terrestrial boreal and tundra ecosystems and of the Greenland and Antarctic ice sheets.

Cases 2' and 3', which impose much more substantial emissions cuts than in cases 2 and 3, yield much more substantial reductions in predicted temperature increase compared with the reference. In addition, significant differences now occur between the two cases (Fig. 3a). In case 3' the temperature increase is reduced by nearly 80% in 2100 and the global-average temperature temporarily stabilizes at ~ 0.5 °C above 1990 levels. At the same time CO₂ levels are ~ 480 p.p.m. (parts per million) and are still rising. In contrast, for case 2' the reduction is only $\sim 55\%$ from the reference by 2100, the temperature is about 1.1 °C above 1990, the CO₂ levels are about 420 p.p.m., and the temperature is still rising. These

differences reflect the fact that GWPs only provide a crude approximation of the effects of these gases on climate, and the errors that show up under our more stringent policy cases have important policy consequences. Use of current 100-year GWPs would mean that a multi-gas policy designed to stabilize the climate (such as that approximated by case 3') may not, in fact, do so, if countries decide to meet the policy target using only reductions in carbon dioxide emissions and carbon sinks. The fact that case 3' shows less warming than case 2' means that other gases as a group are undervalued by 100-year GWPs, compared with the predictions of our model which simulates the principal interactions among gases. If 100-year GWPs were perfect indices then the results from cases 2 and 3 would be identical; and the results from cases 2' and 3' would be identical.

The case of CH₄ and its interactions with other gases illustrates deeper problems with GWPs. For example, as a result of the much higher case 2' emissions (Fig. 2b) in 2100, the atmospheric levels of CH₄, CO and ozone in case 2' are respectively 2.8, 1.8 and 1.2 times those in case 3'. Through chemistry–climate interactions this leads to a 40% lower level in 2100 of the main atmospheric destroyer of CH₄ (the hydroxyl free radical, OH) and proportionately longer CH₄ lifetimes (and, hence, a higher GWP), in case 2' relative to case 3'. Global warming potentials reflect the cumulative forcing over the specified (for example, 100-year) horizon, rather than a radiative forcing effect that is relevant for a particular year. The emissions control path in case 3', with gradually greater reductions in emissions through to 2100, means that the majority of integrated CH₄ emission reductions occur in the few decades approaching 2100. These GWP-weighted reductions in CH₄ are largely reflected as less warming by 2100 because of its shorter lifetime. In contrast, the climate effects of the additional reductions in CO₂ (in case 2' relative to case 3') that occur late in the next century, would be observed mostly over the succeeding century and beyond. In other words, different GWPs are needed for different magnitudes and relative proportions of CO₂ and non-CO₂ reductions and for different times for assessing climatic effects. These complexities defeat the utility of the index.

Ecosystem response

We report ecosystem effects in terms of net primary production (NPP), defined as the difference between plant carbon uptake by photosynthesis and loss by plant respiration, which is an important ecosystem variable. It reflects the ability of ecosystems to supply food and fibre and drives the rates of most “ecosystem services”, including carbon storage^{42,43}. The ability of terrestrial ecosystems to store anthropogenic carbon depends on the relative rates of NPP and the decomposition of organic matter. The five IGSM runs all show NPP increases, but all policy cases show smaller increases (Fig. 3b). Decomposition also increases in all scenarios, but at a slower rate than NPP, such that 161 GtC is added to natural ecosystems between 1990 and 2100 in the reference case. In the model runs simulating the CO₂-only approach (case 2) and the multi-gas approach (case 3), 127 and 130 GtC respectively are added. Cases 2' and 3' show considerably smaller increases in NPP and greater differences occur between the cases. The NPP patterns for case 2' and 3' result from a complex set of interactions between plant and soil responses to changes in both CO₂ concentration and nitrogen availability, as controlled by temperature

changes. These results illustrate an additional limitation of GWPs. In particular, because the fertilization effect of rising CO₂ increases NPP and carbon storage in ecosystems, there is a negative feedback on the climate system. GWPs do not capture this effect and thus the current GWP formulation undervalues all other gases *in toto* relative to CO₂ (or, equivalently, overvalues CO₂ relative to all other gases *in toto*).

The CO₂ fertilization effect compensates for at least some of the climate-induced reductions in plant production and ecosystem storage, until the atmospheric CO₂ concentration reaches a saturation level, after which there would be no additional beneficial effects. Both the magnitude of the effect and the saturation level are uncertain. Some ecological models, including ours, predict that the CO₂ fertilization effect will be large enough to more than compensate for climate-induced reductions in plant production and ecosystem carbon storage throughout the twenty-first century. Other models predict that the climate-induced effects will dominate the ecosystem response and any CO₂ fertilization compensation will disappear during the first half of the twenty-first century⁴⁴.

Our analysis does not include the effects of the potential redistribution of ecosystem boundaries on terrestrial carbon dynamics or climate. Such redistribution shows up in extensive studies of species response to climate changes in the past. Both the speed of the redistribution and the areal extent of the dieback–reorganization sequence over a specific area are uncertain. One reason for this uncertainty is an incomplete understanding of the response of ecosystems to disturbances such as fires, high winds, ice storms, insects or other pathogens, particularly if there are large synchronous disturbances. In a modelling exercise that included large synchronous disturbances and species redistribution⁴⁵, the estimated loss was as much as 200 Gt C before a new steady-state condition was established and net carbon storage occurred again in terrestrial ecosystems. This estimate probably represents an upper bound to the effect of biome shifts alone in causing a transient loss of carbon from the terrestrial biosphere. The faster the rate of climate change, the greater the likelihood of large transient carbon losses from terrestrial ecosystems due to biome shifts during the transient carbon losses⁴⁶. Such losses would alter predicted CO₂ levels and hence warming, compared to those in our study.

Other consequences of vegetation redistribution may be more subtle. Because boreal forests absorb more solar radiation than tundra does, predicted poleward shifts in the location of the forest–tundra boundary during a period of warming can amplify climate changes by as much as 50% (ref. 47). The amplification of warming associated with the poleward movement of the boreal forest would alter the temperature changes implied in Fig. 3.

Discussion

Much current analysis and policy discussion narrows the climate issue to a debate about carbon emissions from fossil fuels. In contrast, we have analysed climate policy as negotiated under the Kyoto agreement including critical issues like forest sinks and the non-CO₂ greenhouse gases. The analysis also explicitly considers atmospheric interactions among these gases, climate feedbacks, the role of CO and NO_x, and the cooling effect of aerosols.

Economic analyses that leave out other trace gases err in several ways: reference emissions are understated, allowable emissions in the commitment period are too low, and opportunities to reduce emissions of other gases are not considered in abatement options. These effects are partially offsetting, so it is not possible to predict the direction of the error *a priori*. Based on our modelling exercise we find that omitting non-CO₂ gases and sinks leads to an overestimate of the annual cost in 2010 in Annex B regions overall, of ~21%. More important, however, is that achieving approximately the same reduction in warming by control of fossil CO₂ only, ignoring other gases and sinks, could cost over 60% more. Failure to consider other trace gases and sinks also has differential regional

effects that could affect the value of emissions trading. A striking effect in our estimates is that failure to reduce leakage from natural gas and coal mining operations could reduce ‘hot air’ in the FSU substantially. Further, the OOE becomes one of the lower-cost regions, whereas it is in fact high-cost if only carbon from fossil fuels is considered. Given the very few studies of the cost of control of non-CO₂ gases, it is not possible to provide a quantitative estimate of probability ranges for these estimates. However, the potential from other gases appears to be substantial, perhaps greater than we have estimated; realizing this potential will require that policy incentives to control these other gases that minimize transactions costs be put in place.

The effects of using GWPs to express non-CO₂ greenhouse-gas emissions in terms of equivalent amounts of CO₂ on atmospheric composition, climate and ecosystems are relatively small if target reductions and participation in the agreement does not change from that negotiated under the Kyoto Protocol. But this is largely due to the relatively small role played by reductions in non-CO₂ gases in this case. Under a much more stringent hypothetical policy that could approximately achieve stabilization of climate, large differences develop between a multi-gas and a CO₂-only strategy. This exposes a significant weakness in the GWP approach. In fact, our predictions show that using Kyoto-specified GWPs means that the choice of control strategy could affect whether or not stabilization of radiative forcing is achieved. These differences develop early in the next century, soon after the point where we suggest that the agreement is extended to developing countries where emissions of (and hence potential reductions in) non-CO₂ greenhouse gases are more substantial.

Throughout we have tried to identify significant uncertainties in our analysis. Our reference scenario should be viewed as one plausible picture of the future rather than a most-likely prediction. The Kyoto policy, although it has been signed, still awaits ratification by the necessary number of countries. Our suggested extensions of policies beyond 2010 were designed to illustrate climate and ecosystem effects, and are not intended to represent economically acceptable, fair, or politically feasible extensions of the Kyoto Protocol. Here we have examined the effects of the addition of other gases and sinks in climate-change policy; we find that these effects could be significant. More research is required in order to understand the role of these gases in mitigating potential climate change. □

Received 14 January; accepted 18 August 1999.

- Adams, R. M., Chang, C., McCarl, B. & Callaway, J. in *Global Change: Economic Issues in Agriculture, Forestry, and Natural Resources* (eds Reilly, J. & Anderson, M.) 273–287 (Westview, Boulder, CO, 1992).
- Richards, K. R. in *Global Change: Economic Issues in Agriculture, Forestry, and Natural Resources* (eds Reilly, J. & Anderson, M.) 288–310 (Westview, Boulder, CO, 1992).
- Cook, E. *Lifetime Commitments: Why Policy-makers Can't Afford to Overlook Fully Fluorinated Compounds* (World Resources Institute, Washington DC, 1995).
- Harnisch, J. & Prinn, R. G. Sulfur hexafluoride emissions. *Environ. Sci. Technol.* **4**, 56a (1999).
- Harnisch, J., Sue Wing, I., Jacoby, H. D. & Prinn, R. G. in *Extraction and Processing Division Congr. 1999* (ed. Mishra, B.) 797–815 (The Minerals, Metals and Materials Soc., Warrendale, PA, 1999).
- Victor, D. G. & MacDonald, G. J. *Future Emissions of Sulfur Hexafluoride and Perfluorocarbons: Implications for Global Policy and Verifying Compliance with the Kyoto Protocol* (International Institute for Applied Systems Analysis, Laxenburg, 1998).
- Victor, D. G. & MacDonald, G. J. A model for estimating future emissions of sulfur hexafluoride and perfluorocarbons. *Clim. Change* (in the press).
- Hoffert, M. I. *et al.* Energy implications of future stabilization of atmospheric CO₂ content. *Nature* **395**, 881–884 (1998).
- Wigley, T. M., Richels, R. & Edmonds, J. A. Economic and environmental choices in the stabilization of climate. *Nature* **379**, 240–243 (1996).
- Nordhaus, W. D. *Managing the Global Commons* 49–74 (MIT Press, Cambridge, MA, 1994).
- Hourcade, J. C. *et al.* in *Climate Change 1995: Economic and Social Dimensions of Climate Change* (eds Bruce, J. P. *et al.*) 297–366 (Cambridge Univ. Press, 1996).
- Prinn, R. *et al.* Integrated global system model for climate policy assessment: feedbacks and sensitivity studies. *Clim. Change* **41**, 469–546 (1999).
- Reilly, J. *et al.* *Multi-Gas Assessment of the Kyoto Protocol* (MIT Joint Program on the Science and Policy of Global Change, Report No. 45, MIT, Cambridge, MA, 1999).
- Yang, Z. *et al.* *The MIT Emissions Prediction and Policy Analysis (EPPA) Model* (MIT Joint Program on the Science and Policy of Global Change, Report No. 6, MIT, Cambridge, MA, 1996).

15. Wang, C., Prinn, R. G. & Sokolov, A. A global interactive chemistry climate model: formulation and testing. *J. Geophys. Res.* **103**, 3399–3417 (1998).
16. Sokolov, A. P. & Stone, P. H. A flexible climate model for use in integrated assessment. *Clim. Dyn.* **14**, 291–303 (1998).
17. Xiao, X. *et al.* Transient climate change and net ecosystem production of the terrestrial biosphere. *Glob. Biogeochem. Cycles* **12**, 345–360 (1998).
18. Schimel, D. *et al.* in *Climate Change 1995: the Science of Climate Change* (eds Houghton, J. T. *et al.*) 65–131 (Cambridge Univ. Press, 1996).
19. Fan, S. *et al.* A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models. *Science* **282**, 442–446 (1998).
20. Schneider, S. H. Detecting climatic change signals: are there any fingerprints? *Science* **263**, 341–347 (1994).
21. Eckaus, R. Comparing the effects of greenhouse gas emissions on global warming. *Energy J.* **13**, 25–34 (1992).
22. Reilly, J. & Richards, K. Climate change damage and the trace gas index issue. *Environ. Resour. Econ.* **3**, 41–61 (1993).
23. Schmalensee, R. Comparing greenhouse gases for policy purposes. *Energy J.* **14**, 245–255 (1993).
24. Reilly, J. in *Economics and Policy Issues in Climate Change* (ed. Nordhaus, W. D.) 243–256 (Resources for the Future, Washington DC, 1998).
25. Lind, R. & Schuler, R. in *Economics and Policy Issues in Climate Change* (ed. Nordhaus, W. D.) 59–96 (Resources for the Future, Washington DC, 1998).
26. Energy Information Administration (EIA) *International Energy Outlook* (Rep. No. DOE/EIA-0484(99)); (<http://www.eia.doe.gov/oiia/ieo99>).
27. Nilsson, S. & Schopfhauser, W. The carbon-sequestration potential of a global afforestation program. *Clim. Change* **30**, 267–293 (1994).
28. Kruger, D. *Integrated Assessment of Global Climate Change: Modeling of Non-CO₂ Gases* 1–17 (Methane and Utilities Branch, US Environmental Protection Agency, Washington DC, 1999).
29. US Environmental Protection Agency, Methane Branch. *Costs of Reducing Methane Emissions in the United States* 1–92 (1999); <http://www.epa.gov/ghginfo>
30. Cole, V. *et al.* in *Climate Change 1995: Impacts, Adaptations, and Mitigation of Climate Change* (eds Watson, R. *et al.*) 726–771 (Cambridge Univ. Press, 1996).
31. Denbaly, M. & Vroomen, H. Dynamic fertilizer nutrient demands for corn: a cointegrated and error-correction system. *Am. J. Agric. Econ.* **75** (1) (1993).
32. Fernandez-Cornejo, J. *Demand and Substitution of Agricultural Inputs in the Central Cornbelt States* (Rep. TB-1816, US Dept of Agriculture, Economic Research Service, Washington, 1993).
33. Oram, D. E., Sturges, W. T., Penkett, S. A., McCulloch, A. & Fraser, P. J. Growth of fluoroform (CHF₃, HFC-23) in the background atmosphere. *Geophys. Res. Lett.* **25**, 35–38 (1998).
34. McCulloch, A. Future consumption and emissions of hydrofluorocarbon (HFC) alternatives to CFCs: comparison of estimates using top-down and bottom-up approaches. *Environ. Int.* **21**, 353–362 (1995).
35. Maiss, M. & Brenninkmeijer, C. A. M. Atmospheric SF₆, trends, sources and prospects. *Environ. Sci. Technol.* **32**, 3077–3086 (1998).
36. de Jager, D., Hendriks, C. A., Heijnes, H. A. M. & Blok, K. in *Greenhouse Gas Control Technologies* (eds Eliasson, B., Riemer, P. W. F. & Wokaun, A.) 503–508 (Pergamon, New York, 1999).
37. Jacoby, H. *et al.* CO₂ limits: economic adjustments and the distribution of burdens. *Energy J.* **18** (3), 31–58 (1997).
38. US DOE Interlaboratory Working Group. *Scenarios of US Carbon Reductions* (Office of Efficiency and Renewable Energy, US Dept of Energy, Washington DC, 1997).
39. Jacoby, H. The uses and misuses of technology development as a component of climate policy 1–16 (MIT Joint Program on the Science and Policy of Global Change, Report No. 43, MIT, Cambridge, MA, 1998).
40. Goulder, L. & Schneider, S. Induced technical change and the attractiveness of CO₂ abatement policies. *Resour. Energy Econ.* **21**, 211–253 (1999).
41. Wang, C. & Prinn, R. G. Impact of emissions, chemistry, and climate on atmospheric carbon monoxide: 100-year predictions from a global chemistry-climate model. *Chemosphere* (in the press); (also as the MIT Joint Program on the Science and Policy of Global Change Report No. 35, MIT, Cambridge, MA, 1998).
42. Daily, G. C. (ed.) *Daily Services: Societal Dependence on Natural Ecosystems* (Island Press, Washington DC, 1997).
43. Melillo, J. M. Warm, warm on the range. *Science* **283**, 183–184 (1999).
44. Cao, M. & Woodward, F. I. Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. *Nature* **393**, 249–252 (1998).
45. Smith, T. M. & Shugart, H. H. The transient response of terrestrial carbon storage to a perturbed climate. *Nature* **361**, 523–526 (1993).
46. Woodwell, G. M. *et al.* in *Biotic Feedbacks in the Global Climatic System* (eds Woodwell, G. M. & Mackenzie, F. T.) 393–411 (Oxford Univ. Press, NY, 1995).
47. Foley, J. A. *et al.* Feedbacks between climate and boreal forests during the mid-Holocene. *Nature* **371**, 52–54 (1994).

Acknowledgements

We acknowledge the financial support of the industrial and government sponsors of the Joint Program on the Science and Policy of Global Change at MIT, and thank V. Webb for his research assistance.

Correspondence should be addressed to J.R. (e-mail: jreilly@mit.edu).

Commitment to the B-lymphoid lineage depends on the transcription factor Pax5

Stephen L. Nutt^{†‡}, Barry Heavey[†], Antonius G. Rolink[‡] & Meinrad Busslinger[†]

[†] Research Institute of Molecular Pathology, Dr Bohr-Gasse 7, A-1030 Vienna, Austria

[‡] Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland

The *Pax5* gene encoding the B-cell-specific activator protein (BSAP) is expressed within the haematopoietic system exclusively in the B-lymphoid lineage, where it is required *in vivo* for progression beyond the pro-B-cell stage. However, Pax5 is not essential for *in vitro* propagation of pro-B cells in the presence of interleukin-7 and stromal cells. Here we show that pro-B cells lacking Pax5 are also incapable of *in vitro* B-cell differentiation unless Pax5 expression is restored by retroviral transduction. *Pax5*^{-/-} pro-B cells are not restricted in their lineage fate, as stimulation with appropriate cytokines induces them to differentiate into functional macrophages, osteoclasts, dendritic cells, granulocytes and natural killer cells. As expected for a clonogenic haematopoietic progenitor with lymphomyeloid developmental potential, the *Pax5*^{-/-} pro-B cell expresses genes of different lineage-affiliated programmes, and restoration of Pax5 activity represses this lineage-promiscuous transcription. Pax5 therefore plays an essential role in B-lineage commitment by suppressing alternative lineage choices.

All types of blood cell are generated from a pluripotent haematopoietic stem cell (HSC) through developmentally restricted progenitors which undergo lineage commitment and subsequent differentiation along a single pathway. The lymphoid lineages develop through a common lymphoid progenitor (CLP) which gives rise to natural killer, B and T cells¹. Early B-cell development can be dissected into different stages according to the rearrangement status of the immunoglobulin heavy-chain (*IgH*) locus, the expression of stage-specific cell-surface markers and growth factor requirements^{2,3}. The earliest B-lineage precursor cells carry the *IgH* locus still in germline configuration. *D_H-J_H* recombination is subsequently initiated in pre-BI cells³, which are also known as early pro-B cells (fraction B)². These pro-B cells can be cultured *in vitro* on stromal cells in the presence of interleukin-7 (IL-7) and express the B-cell surface proteins λ5, VpreB, Igα and Igβ (refs 3, 4). Completion of a functional *V_H-D_H-J_H* rearrangement results in the expression of the pre-B-cell receptor and subsequent differentiation to small pre-B cells, which are no longer responsive to pro-B-cell growth conditions⁵.

The initiation of B-cell development critically depends on two transcription factors; the basic helix-loop-helix proteins encoded by the *E2A* gene and the early B-cell factor (EBF). In the absence of either protein, B-cell development is aborted at the earliest stage, before *D_H-J_H* rearrangement of the *IgH* gene⁶⁻⁸. Moreover, forced expression of E2A and EBF in haematopoietic precursor cells revealed that these regulators cooperatively induce the transcription of several B-lymphoid-specific genes^{9,10}. Hence, loss- and gain-of-function experiments have implicated E2A and EBF in the control of B-lineage commitment.

A third transcriptional regulator involved in early B-lymphopoiesis is the B-cell-specific activator protein (BSAP), which is encoded by the *Pax5* gene and recognizes DNA through the highly conserved paired domain characteristic of the Pax family of transcription factors¹¹. Within the haematopoietic system, *Pax5* is exclusively expressed in the B-lymphoid lineage from the earliest detectable precursor to the mature B-cell stage^{12,13}. In the absence of Pax5, B-cell development is arrested in the fetal liver at a similarly early stage as in *E2A* and *EBF* mutant mice, whereas it proceeds in adult bone marrow up to the early pro-B (pre-BI) cell stage^{14,15}. B-lymphopoiesis in the bone marrow is, however, not advanced by expression of a functionally rearranged μ-heavy-chain transgene⁵, indicating that

the observed developmental block does not result from the inability of *Pax5*^{-/-} pro-B cells to undergo *V_H-D_H-J_H* rearrangements at the *IgH* locus¹⁵. *Pax5*^{-/-} pro-B cells can be cultured *in vitro* with similar efficiency as wild-type pro-B cells, a fact that has facilitated the genetic identification of Pax5 target genes¹⁵. BSAP (Pax5) functions both as a transcriptional activator and as a repressor, as it positively controls *CD19*, *mb-1* (*Igα*), *N-myc* and *Lef-1* expression and negatively regulates *PD-1* transcription¹⁶. Apart from these target genes, Pax5-deficient and wild-type pro-B cells do not significantly differ in the expression of 50 B-cell-associated genes analysed¹⁶. This observation suggested that pro-B cells, even in the absence of Pax5, still undergo commitment to the B-lymphoid lineage. However, investigating the development potential of *Pax5*^{-/-} pro-B cells, we now show that Pax5, instead of E2A and EBF, is the critical transcription factor involved in B-lineage commitment.

Pax5^{-/-} pro-B cells lack B-lymphoid potential

The *RAG2* mutation also blocks B-cell development at the pro-B-cell stage *in vivo*¹⁷, like the *Pax5* mutation¹⁵, but it is compatible with

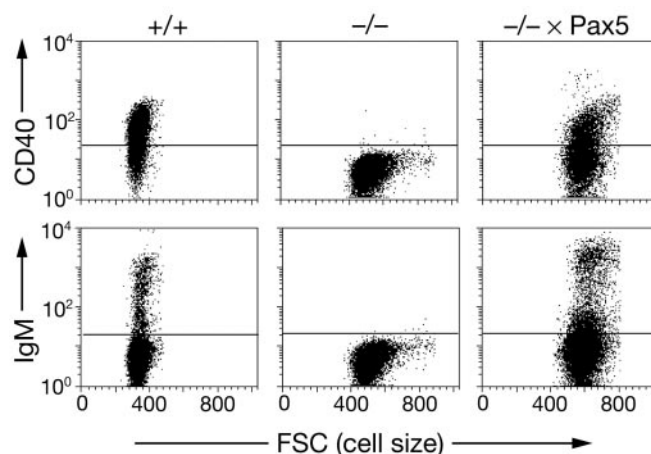


Figure 1 Pax5 is essential for *in vitro* differentiation of B lymphocytes. Wild-type and *Pax5*^{-/-} pro-B cells expressing a retroviral *bcl2* gene were cultured in the absence of stromal cells and IL-7 for three days in IMDM medium alone followed by flow cytometric analysis. Gating for CD19⁺ cells was used to select the *Pax5*^{-/-} pro-B cells infected with a BSAP (Pax5)-expressing retrovirus¹⁶. FSC, forward scatter.

* Present address: Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, UK.

differentiation of these cells to mature B-lymphocytes *in vitro*¹⁸. We have therefore examined the *in vitro* differentiation potential of *Pax5*^{-/-} pro-B cells by culturing them for three days without IL-7 and stromal (ST2) cells. To promote cell survival upon IL-7 withdrawal¹⁸, a retroviral *bcl2* gene was introduced into *Pax5*^{-/-} and wild-type pro-B cells. In the absence of IL-7 and ST2 cells, the wild-type cells complete the rearrangement of their immunoglobulin genes and differentiate to small CD40⁺IgM⁺ B cells¹⁸ (Fig. 1). In contrast, *Pax5*^{-/-} cells do not express any markers associated with late differentiation (Fig. 1; and data not shown). However, *Pax5*^{-/-} pro-B cells infected with a Pax5-expressing retrovirus can differentiate into B cells *in vitro* (Fig. 1). Hence, the *Pax5* mutation arrests B-cell development at the same early stage *in vitro* as *in vivo*, and this developmental block can be overcome by retrovirus-mediated expression of *Pax5*.

Pax5^{-/-} pro-B cells are uncommitted

While growing pro-B-cell clones, we found that *Pax5*^{-/-} pro-B cells, in contrast to wild-type pro-B cells, could survive for up to three weeks without medium change. Although the culture medium was depleted of IL-7 during this time, some *Pax5*^{-/-} cells underwent a characteristic change in morphology and subsequently continued to proliferate. Whereas *Pax5*^{-/-} pro-B cells grown in IL-7 are small and refractile, and contain little cytoplasm (Fig. 2a), these cells became larger and irregular in shape upon IL-7 depletion (Fig. 2b).

To systematically investigate this phenomenon, we sorted individual pro-B cells into single wells of a 96-well plate, scored the growth of pro-B-cell colonies after 10 days in IL-7 medium and then monitored the frequency with which these colonies changed their morphology as IL-7 became limiting (Fig. 2c). For this purpose, pro-B cells were either freshly isolated from the bone marrow of *Pax5*^{-/-}, *RAG2*^{-/-} and wild-type mice (*ex vivo*) or were sorted from established pro-B cell cultures (*in vitro*). The *Pax5*^{-/-} pro-B-cell clones changed their morphology at a high frequency (64–87%), irrespective of *bcl2* expression, whereas the wild-type and *RAG2*^{-/-} pro-B cells did not show this morphology change (Fig. 2c). Our interpretation of these results is that the wild-type and *RAG2*^{-/-} pro-B cells have undergone B-lineage commitment and can therefore only differentiate to non-proliferating B-cells upon IL-7

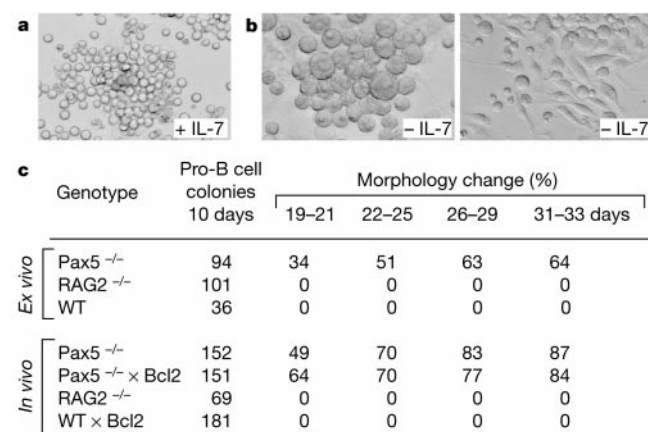


Figure 2 *Pax5*-deficient pro-B cells are not restricted to the B-lymphoid lineage. **a**, Confocal microscopic image of a *Pax5*^{-/-} pro-B-cell colony after 10 days in IL-7 medium. **b**, Altered morphology of two representative colonies of the same pro-B-cell culture 30 days after the last medium change. **c**, Statistical analysis. Individual B220⁺ c-Kit⁺ pro-B cells, which were isolated by single-cell sorting directly from the bone marrow (*ex vivo*) or from established pro-B-cell cultures (*in vitro*) of the indicated genotypes, were grown on stromal ST2 cells in IL-7 medium, scored for colony formation at day 10 and subsequently examined for altered morphology at 3-day intervals (without medium change). The *bcl2* gene was introduced by retroviral transduction. WT, wild-type.

withdrawal¹⁸ (Fig. 1). However, under the same conditions *Pax5*^{-/-} pro-B cells cannot develop along the B-lymphoid lineage (Fig. 1). Instead, they appear to differentiate along other haematopoietic pathways, indicating that these cells are not committed to the B-lymphoid lineage.

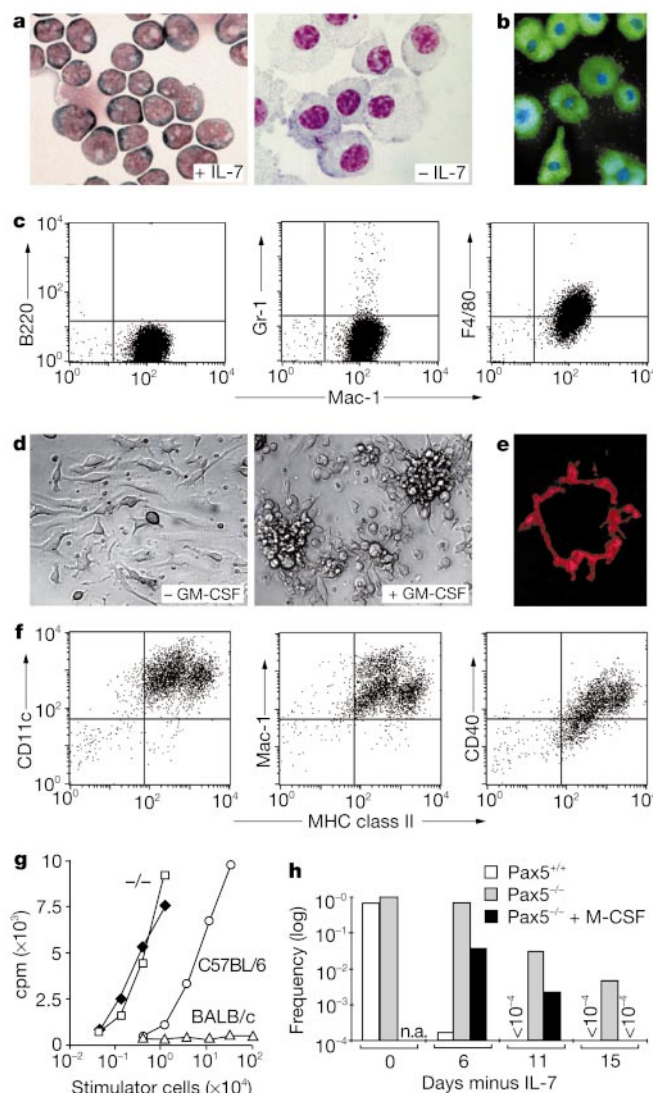


Figure 3 *In vitro* generation of macrophages and dendritic cells. **a**, May–Grünwald–Giemsa staining of cytopsin preparations of *Pax5*^{-/-} pro-B cells cultured for 14 days on ST2 cells with or without IL-7. **b**, Phagocytosis. *In vitro* differentiated *Pax5*^{-/-} macrophages were incubated with FITC-labelled *E. coli*. **c**, Flow cytometric analysis of macrophages derived from *Pax5*^{-/-} pro-B cells. **d**, Confocal microscopic images of *Pax5*^{-/-} pro-B cells grown for 16 days on ST2 cells with or without GM-CSF. **e**, MHC class II expression on cells differentiated in GM-CSF medium and stained with an anti-MHC class II antibody. **f**, Flow cytometric analysis of dendritic cells derived from *Pax5*^{-/-} pro-B cells. The cells shown in **c** and **f** were gated on large size. **g**, Mixed leukocyte reaction. Increasing doses of γ -irradiated stimulator cells (cells per well) were incubated for four days with 2.5×10^5 responder cells isolated from lymph nodes of BALB/c mice. [³H]thymidine incorporation (c.p.m.) was measured during the last 14 h of incubation. Dendritic cells derived from two *Pax5*^{-/-} pro-B-cell lines (squares, diamonds) and splenocytes from allogeneic C57BL/6 (circles) or syngeneic BALB/c mice (triangles) were used as stimulator cells. **h**, Kinetic analysis of myeloid commitment. *Bcl2*-expressing pro-B cells deprived of IL-7 and ST2 cells were cultured with or without M-CSF. Viable cells were assessed for their ability to be re-cloned as pro-B cells in the presence of IL-7 and ST2 cells by limiting dilution analysis⁴. The re-cloning frequency is indicated relative to that of *Pax5*^{-/-} pro-B cells at day 0. n.a., not applicable.

In vitro differentiation along monocytic lineages

The morphology of the cells shown in Fig. 2b was suggestive of myeloid cell types, in agreement with the fact that ST2 cells produce the cytokine macrophage colony-stimulating factor (M-CSF)¹⁹. We therefore assessed the potential of *Pax5*^{-/-} pro-B cells to develop into macrophages by culturing them for 10–14 days on ST2 cells (in the absence of IL-7) followed by propagation in M-CSF (with or without ST2 cells). Cells differentiated in this manner displayed the characteristic morphology of enlarged vacuolar macrophages (Fig. 3a) and were positive for the macrophage markers Mac-1 and F4/80, but did not express the B-cell protein B220 and granulocyte marker Gr-1 (Fig. 3c). The same cells efficiently phagocytosed fluorescently labelled *Escherichia coli* (Fig. 3b), indicating that *Pax5*^{-/-} pro-B cells can differentiate into mature macrophages.

Using a cell-cloning assay, we next investigated the kinetics with which *Pax5*^{-/-} pro-B cells become committed to the myeloid lineage. *Bcl2*-expressing wild-type and *Pax5*^{-/-} pro-B cells were cultured in the absence of IL-7 and ST2 cells or in the presence of M-CSF for defined times. Subsequently, the proportion of viable cells that could be re-cloned as pro-B cells upon re-addition of stromal cells and IL-7 was determined by limiting dilution analysis. Wild-type pro-B cells rapidly lost their ability to be re-cloned (Fig. 3h) owing to efficient differentiation along the B-lymphoid lineage⁴. In contrast, the *Pax5*^{-/-} pro-B cells, which are not yet committed to a lineage-specific programme, could be re-cloned even after 15 days of IL-7 and stromal cell withdrawal (Fig. 3h). M-CSF addition, however, reduced the cloning frequency of *Pax5*^{-/-} pro-B cells 10-fold after six days and completely abolished it after 15 days (Fig. 3h). Hence, most cells initiated differentiation along the myeloid lineage and thus lost their IL-7 responsiveness within six days of M-CSF stimulation.

Dendritic cells also develop in the bone marrow from the same monocyte precursor as the macrophages, but require the cytokine granulocyte-macrophage CSF (GM-CSF) instead of M-CSF for terminal differentiation^{20,21}. To generate dendritic cells from *Pax5*^{-/-} pro-B cells, we applied a similar strategy as for macrophage development except that GM-CSF was added in the final differentiation step. After 16 days of GM-CSF stimulation, the cells were prolifer-

ating in small aggregates characteristic of dendritic cells²⁰, whereas only macrophages grew in the absence of GM-CSF (Fig. 3d). Mature dendritic cells are antigen-presenting cells characterized by a stellate shape with multiple cytoplasmic projections and by high levels of MHC class II protein on their surfaces^{20,21}. The *Pax5*^{-/-} cells, differentiated in response to GM-CSF, exhibited the characteristic cell morphology and MHC class II expression (Fig. 3e) as well as the expected cell-surface phenotype (MHC-II^{high} CD11c⁺ Mac-1⁺ CD40⁺ CD8⁻ B220⁻) of mature dendritic cells (Fig. 3f; and data not shown). These cells were further identified as dendritic cells by their ability to stimulate T-cell proliferation in a mixed leukocyte reaction (Fig. 3g). In this assay, allogeneic splenocytes of C57BL/6 mice elicit a proliferative response in T cells of BALB/c mice, in contrast to syngeneic splenocytes (Fig. 3g). Importantly, the *Pax5*^{-/-} dendritic cells activated the T cells with a 10–30-fold higher efficiency than allogeneic splenocytes (Fig. 3g), in agreement with the fact that dendritic cells are significantly more potent than splenocytes in stimulating T-cell proliferation²¹. *Pax5*^{-/-} pro-B cells can therefore also differentiate to become functional dendritic cells.

Osteoclasts constitute a third lineage which originates from a common monocyte precursor. Differentiation to bone-resorbing osteoclasts is controlled by the cytokine TRANCE (also known as OPGL, ODF or RANKL), which is presented on activated osteoblasts²². Following IL-7 withdrawal, we therefore cultured *Pax5*^{-/-} pro-B cells on stromal ST2 cells which ectopically expressed TRANCE. Under these conditions, the pro-B cells differentiated within two weeks into cells expressing the osteoclast-specific enzyme tartrate-resistant acid phosphatase (TRAP; Fig. 4a). These differentiated cells had the multinucleated morphology of mature osteoclasts (Fig. 4b) and formed resorption lacunae on artificial bone substrates (Fig. 4c). Hence, *Pax5*^{-/-} pro-B cells can also develop into functional osteoclasts.

Differentiation into granulocytes

Myeloid progenitors also give rise to cell types of the neutrophil lineage which characteristically contain a highly segmented nucleus and differentiate under the influence of the cytokines IL-6 and granulocyte CSF (G-CSF)²³. To test the granulocytic potential of *Pax5*^{-/-} pro-B cells, we cultured these cells for three weeks with the multilineage cytokines IL-3, IL-6 and stem-cell factor (SCF). Under these conditions, a small percentage of the *Pax5*^{-/-} cells differentiated into granulocytic cells, as shown by their segmented nuclear morphology (Fig. 4e). Homogeneous differentiation to granulocytes was achieved by replacing the multilineage cytokines with G-CSF for six days (Fig. 4f). *Pax5*^{-/-} pro-B cells can, therefore, also differentiate along the granulocytic lineage.

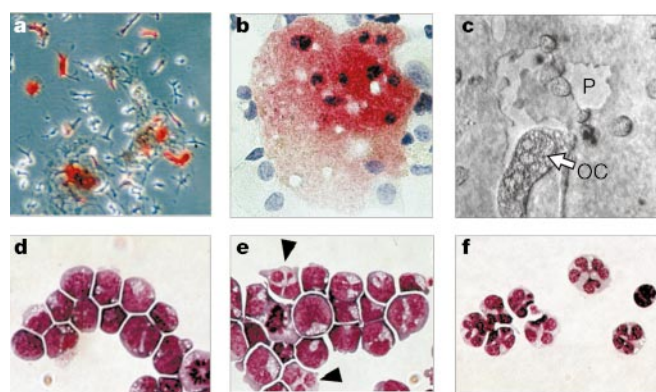


Figure 4 *In vitro* differentiation of osteoclasts and granulocytes. **a**, TRAP staining (red) of *Pax5*^{-/-} cells cultured for two weeks on TRANCE-expressing ST2 cells. **b**, Multinucleated morphology of TRAP⁺ cells. Differentiated *Pax5*^{-/-} cells were stained for TRAP activity (red) followed by counterstaining with haematoxylin (revealing the nuclei in blue) and cytopsin centrifugation. **c**, Bone resorption. Osteoclasts derived from *Pax5*^{-/-} pro-B cells were incubated on an osteologic bone disc with ST2-TRANCE cells. A multinucleated osteoclast (OC) is shown with the pits (P) resulting from its bone-resorbing activity. **d–f**, Cytopsin preparations of May–Grünwald–Giemsa-stained cells. *Pax5*^{-/-} pro-B cells initially grown in IL-7 medium (**d**) were cultured in the presence of IL-3, IL-6 and SCF for three weeks (**e**) and subsequently in G-CSF medium alone for six days (**f**). Granulocytes were identified by their segmented nuclear morphology (arrowhead in **e**).

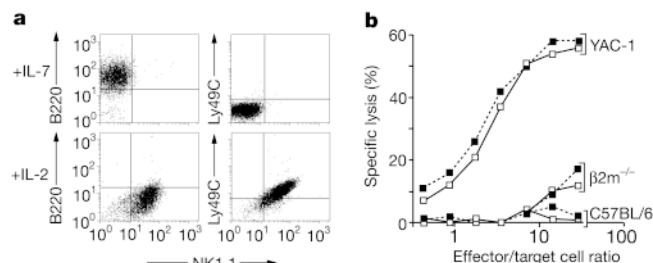


Figure 5 Generation of mature NK cells. **a**, Flow cytometric analysis of *Pax5*^{-/-} pro-B cells cultured for nine days with IL-2 and ST2 cells. **b**, Cytotoxicity. Cells of two *Pax5*^{-/-} pro-B-cell lines (closed and open symbols) were differentiated in IL-2 for nine days and assayed for their cytotoxicity against YAC-1 cells and LPS-treated splenocytes from wild-type C57BL/6 or $\beta_2m^{-/-}$ mice²⁷. Cytotoxicity assays were performed at the indicated effector to target cell ratios.

Differentiation into natural killer cells

The myeloid differentiation experiments unequivocally show that *Pax5*^{-/-} pro-B cells are not committed to the B-lymphoid lineage. Hence, the question arose of whether these cells also have properties of the recently described CLP and could thus give rise to natural killer (NK) cells¹. NK-cell development critically depends on IL-15 (ref. 24), which can, at least *in vitro*, be substituted by IL-2 (ref. 25). To assess their NK-cell potential, the *Pax5*^{-/-} pro-B cells were therefore cultured with IL-2 and stromal cells. Within 10 days, the *Pax5*^{-/-} cells switched off expression of the B-cell protein B220, initiated synthesis of the NK-cell markers NK1.1, Ly49C and 2B4 and were identified as NK cells by their cell-surface phenotype (NK1.1⁺ Ly49C⁺ 2B4⁺ IL-2Rα^{low} IL-2Rβ^{low} Mac-1^{low} CD3ε⁻ CD4⁻ CD8⁻ B220⁻; Fig. 5a and data not shown).

NK cells contribute to innate immunity by killing host cells that fail to express MHC class I proteins on their surfaces as a consequence of viral infection or tumorigenesis²⁶. We therefore measured the cytolytic activity of *in vitro* differentiated *Pax5*^{-/-} cells in a ⁵¹Cr release assay, using the sensitive YAC-1 tumour cells or LPS-stimulated lymphoblasts from β₂-microglobulin-deficient (β₂m^{-/-}) mice, which lack MHC class I cell-surface expression²⁷. Both target cells were lysed in a dose-dependent manner by NK cells derived from *Pax5*^{-/-} pro-B cells, whereas LPS-activated target cells from a syngeneic C57BL/6 mouse were resistant to lysis (Fig. 5b). Hence, *Pax5*^{-/-} pro-B cells can differentiate into functional NK cells.

Under all *in vitro* differentiation conditions analysed, we never detected *Pax5*^{-/-} cells that expressed surface antigens characteristic of the T-lymphoid lineage, in agreement with the fact that the complex inductive microenvironment of the thymus is required for early T-cell development. When injected into *RAG2*^{-/-} mice, *Pax5*-deficient pro-B cells were, however, able to fully reconstitute T-cell development *in vivo*²⁸. Therefore, *Pax5*^{-/-} pro-B cells can also give rise to lymphocytes except for B cells.

In vivo myeloid development of *Pax5*^{-/-} pro-B cells

The observation that injected *Pax5*^{-/-} pro-B cells home to the bone

marrow and undergo self-renewal in *RAG2*^{-/-} mice²⁸ raises the question of whether these cells can also differentiate *in vivo* into myeloid cells, in addition to T-lymphocytes. However, we could not detect the formation of myeloid cell types in reconstituted *RAG2*^{-/-} mice²⁸, suggesting that the *Pax5*^{-/-} pro-B cells cannot efficiently compete with endogenous myeloid progenitors. For this reason, we analysed the *in vivo* myeloid potential of *Pax5*^{-/-} pro-B cells in the *c-fos*^{-/-} mouse, which has a developmental defect in the osteoclast lineage^{29,30}. Consequently, *c-fos*^{-/-} mice develop osteopetrotic bones and, owing to the absence of a bone marrow cavity, exhibit extramedullary haematopoiesis in the spleen³⁰. As transplantation of *Pax5*^{-/-} pro-B cells alone cannot rescue lethally irradiated mice (data not shown), we performed competitive reconstitution experiments by injecting *Pax5*^{-/-} pro-B cells together with *c-fos*^{-/-} splenocytes into lethally irradiated, newborn *c-fos*^{-/-} mice. Four weeks after cell transfer, osteoclasts derived from *Pax5*^{-/-} pro-B cells could be detected by TRAP staining on the surface of the calvariae of three reconstituted *c-fos*^{-/-} mice (Fig. 6c), as in wild-type mice (Fig. 6a), whereas calvarial osteoclasts were never found in untreated *c-fos*^{-/-} mice (Fig. 6b). However, the reconstituted *c-fos*^{-/-} mice showed neither tooth eruption nor development of a bone marrow cavity in the long bones (Fig. 6f). A few clusters of TRAP⁺ osteoclasts were nevertheless observed in the osteopetrotic bones of the reconstituted *c-fos*^{-/-} mice (Fig. 6f), in contrast to untreated *c-fos*^{-/-} mice (Fig. 6e). Hence, the injected *Pax5*^{-/-} pro-B cells gave rise to only a partial rescue of the *c-fos* mutant phenotype, possible owing to still inefficient competition with myeloid progenitors derived from the *c-fos*^{-/-} HSC. Nevertheless, these experiments show that *Pax5*^{-/-} pro-B cells can differentiate into at least one myeloid cell type *in vivo*.

The multilineage potential is clonal

During this study, we tested the myeloid and lymphoid differentiation potential of five independently derived *Pax5*^{-/-} pro-B-cell pools, of which three were infected with a *bcl2*-expressing retrovirus. All five pro-B-cell lines developed into macrophages *in vitro* and T cells

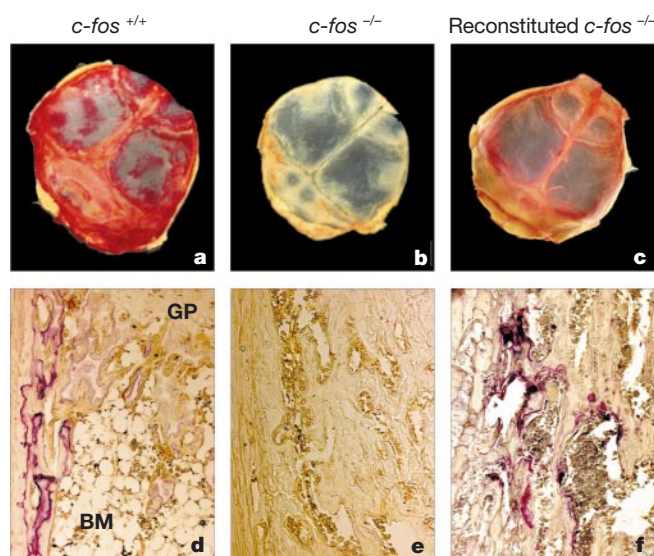


Figure 6 Development of osteoclasts from *Pax5*^{-/-} pro-B cells *in vivo*. Identification of osteoclasts in the calvariae (a–c) and long bones (d–f) of 4-week-old mice. TRAP staining (red) was used to detect osteoclasts on the calvarial surface and in sections of the tibia of wild-type (a, b), *c-fos*^{-/-} (b, e) and reconstituted *c-fos*^{-/-} (c, f) mice. For reconstitution, 10⁷ *Pax5*^{-/-} pro-B cells and 10⁷ splenocytes, isolated from an adult *c-fos*^{-/-} mouse, were co-injected intraperitoneally into lethally irradiated *c-fos*^{-/-} mice one day after birth. Four weeks later, osteoclast formation was analysed by TRAP histochemistry. BM, bone marrow cavity; GP, growth plate.

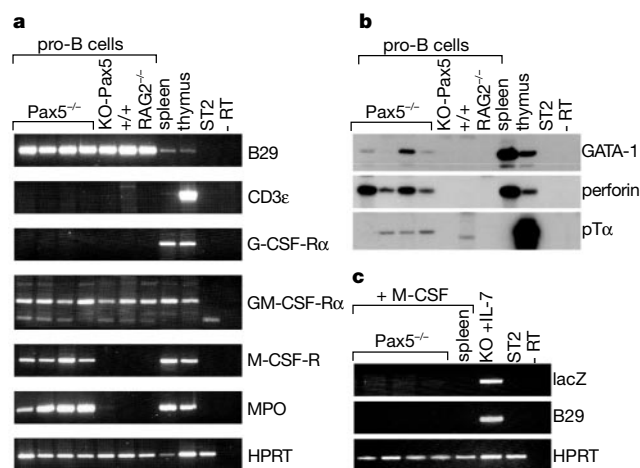


Figure 7 Lineage-promiscuous gene expression in *Pax5*^{-/-} pro-B cells. **a**, **b**, RT-PCR analysis. The following pro-B cells grown on stromal ST2 cells in the presence of IL-7 were analysed: four independent *Pax5*^{-/-} pro-B cell lines, *Pax5*^{-/-} pro-B cells reconstituted with a *Pax5*-expressing retrovirus (KO-*Pax5*) and pro-B cells established from wild-type (+/+) and *RAG2*^{-/-} bone marrow. The PCR products were visualized on agarose gels by ethidium bromide staining (**a**) or Southern blot analysis (**b**). The quantity of input cDNA was normalized by analysing the control *HPRT* transcript. No reverse transcriptase was added in lane – RT. **c**, Loss of *Pax5* expression upon myeloid differentiation. *Pax5*^{-/-} pro-B cells and wild-type splenocytes were cultured for 10 days with M-CSF and ST2 cells. *Pax5* expression was subsequently monitored by RT-PCR with primers from *Pax5* exon 1A and *lacZ* sequences inserted in the targeted *Pax5* locus¹⁴. *Pax5*^{-/-} pro-B cells grown in IL-7 medium (KO + IL-7) were used as positive control.

in vivo, indicating that the ability to differentiate along the myeloid and lymphoid lineages is an intrinsic property of all *Pax5*^{-/-} pro-B cells, irrespective of *bcl2* expression. However, it could be argued that, despite their growth in IL-7, the *Pax5*^{-/-} pro-B-cell pools may correspond to a heterogeneous mixture of lineage-committed progenitors. To exclude this possibility, we have analysed the differentiation potential of eight subclones which were generated by single-cell sorting of *Pax5*^{-/-} pro-B-cell pools. In the appropriate environment, all eight subclones efficiently developed into macrophages, osteoclasts, dendritic cells (*in vitro*) and T cells (*in vivo*). The differentiation into granulocytes and NK cells was variable, possibly indicating either that our *in vitro* differentiation conditions were not optimal for these two lineages or that *in vitro* cultured *Pax5*^{-/-} pro-B cells gradually lose the capacity to generate these two cell types. Nevertheless, our results unequivocally show that *Pax5*^{-/-} pro-B cells are multipotent haematopoietic progenitors which maintain their differentiation potential after single-cell cloning.

Pax5 represses lineage-promiscuous transcription

Multipotent haematopoietic progenitors simultaneously express genes from different lineage-affiliated programs, in a process known as 'multilineage priming'³¹. Consistent with its progenitor status, the *Pax5*^{-/-} pro-B cell is therefore expected to exhibit lineage-promiscuous gene expression, even though *Pax5*^{-/-} pro-B cells differ only minimally from committed wild-type pro-B cells in the expression of 50 B-cell-associated genes analysed¹⁶. To investigate this issue, we used the polymerase chain reaction with reverse transcription (RT-PCR) to compare the expression of different lineage-specific genes between uncommitted *Pax5*^{-/-} pro-B-cell lines and committed pro-B cells derived either by retrovirus-mediated reconstitution of *Pax5* expression or by isolation from wild-type and *RAG2*^{-/-} bone marrow (Fig. 7a, b). According to their expression pattern, the different genes could be grouped into three categories (Fig. 7; and data not shown). The transcripts of several genes, including *CD3ε* (T-lymphoid), *G-CSF-Rα* (granulocytic) and *β^{maj}-globin* (erythroid), were not detected in any pro-B-cell line, whereas they could be readily amplified from control tissues. Other genes were equally expressed in pro-B cells of all genotypes. This category comprises the genes *B29* (B-lymphoid), *GM-CSF-Rα* (myeloid), *EPO-R* (erythroid), *Tcf-1* (T-lymphoid) and *GATA-3* (T-lymphoid), although the last two transcription factor genes were expressed only at a very low level. Some genes, on the other hand, were exclusively expressed in *Pax5*^{-/-} pro-B cells, but not in any of the committed pro-B cells. This group includes the relatively abundant transcripts of the myeloid M-CSF receptor (*M-CSF-R*; *c-fms*) and myeloperoxidase (*MPO*) genes as well as the rare messenger RNAs of the *GATA-1* (erythroid), *perforin* (NK cell) and *pTα* (T lymphoid) genes. Importantly, the lineage-promiscuous expression of these genes was efficiently repressed in *Pax5*^{-/-} pro-B cells by restoring *Pax5* activity using retroviral transduction (Fig. 7a, b).

We next investigated whether the transcription of B-cell-specific genes such as *B29* (*Igβ*) and *Pax5* is repressed in *Pax5*^{-/-} pro-B cells upon entry into the myeloid lineage. Ten days after replacement of IL-7 with M-CSF, four independent *Pax5*^{-/-} cell lines efficiently downregulated the expression of *B29* as well as of the *lacZ* gene inserted in the targeted *Pax5* locus (Fig. 7c). In summary, our data indicate that an essential function of *Pax5* in B-lineage commitment is to repress the lineage-promiscuous transcription of non-B-lymphoid genes. Conversely, transcription of the *Pax5* gene is itself inactivated upon differentiation to other haematopoietic lineages.

Discussion

Differentiation of the HSC into distinct blood cell types is thought to progress through intermediate progenitor cells with restricted developmental potential. This view of haematopoiesis is challenged

by our finding that the *Pax5*^{-/-} pro-B cell possesses, at least under *in vitro* conditions, a broad developmental potential similar to that of the HSC itself. So far, however, we have been unable to differentiate the *Pax5*^{-/-} pro-B cell along the erythroid and megakaryocytic lineages, which may reflect a more restricted developmental capacity of this cell or our failure to define appropriate differentiation conditions for these lineages. Moreover, *Pax5*^{-/-} pro-B cells have so far failed to reconstitute the entire haematopoietic system in transplantation experiments²⁸. In the absence of this characteristic property of the HSC, the *Pax5*^{-/-} pro-B cell must be classified as a haematopoietic progenitor cell with broad lymphoid and myeloid differentiation potential.

A close developmental connection between the lymphoid and myeloid lineages has been demonstrated by the identification of a B-cell/macrophage-restricted precursor, on the basis of several criteria³². First, certain human acute leukaemias co-express B-cell and macrophage characteristics, indicating that they may be derived from a transformed progenitor cell of both lineages³³. Second, mice expressing the *Eμ-myc* transgene generate immature lymphomas that can differentiate along the B-lymphoid and macrophage lineages^{34,35}. Third, several immortalized pre-B-cell lines can be converted into macrophages³⁶⁻³⁸. Finally, the murine fetal liver contains a small subpopulation of clonogenic B-cell/macrophage progenitors³⁹. Fetal liver cells with an expanded lymphomyeloid potential have been identified by several groups by varying the conditions of their *in vitro* differentiation assays⁴⁰⁻⁴². A progenitor cell with a similar developmental capacity has so far not been isolated from mouse bone marrow¹, whereas human bone marrow contains cells with the potential to differentiate along the T, B, NK and dendritic cell lineages⁴³. The interpretation of these studies is, however, complicated by the fact that the isolated progenitor cells are rare, correspond to only a transition stage in haematopoietic development and cannot be expanded long-term *in vitro* in an uncommitted state, thus precluding any detailed analysis. In contrast, the *Pax5*^{-/-} pro-B cell can be propagated *in vitro* as an uncommitted haematopoietic progenitor. Moreover, this cell originates from the bone marrow and, except for B-cell development, exhibits the full spectrum of myeloid and lymphoid differentiation.

The development of lymphocytes, but not of myeloid cells, is critically dependent on IL-7, which promotes the proliferation and/or survival of lymphoid precursors⁴⁴. Expression of the IL-7 receptor was used as a criterion for isolating the CLP from mouse bone marrow¹. This CLP has the capacity to rapidly reconstitute B, T and NK cells *in vivo*, but lacks myeloid differentiation potential *in vivo* and *in vitro*¹. A close relationship between the CLP and *Pax5*^{-/-} pro-B cell is suggested by the fact that the *Pax5*^{-/-} pro-B cell also expresses the IL-7 receptor¹⁵, depends on IL-7 for *in vitro* propagation¹⁵ and possesses lymphoid developmental potential. However, in contrast to the CLP, the *Pax5*^{-/-} pro-B cell can differentiate along myeloid lineages *in vitro* and generate osteoclasts *in vivo* after injection into *c-fos*^{-/-} mice. Given the transitory nature of the CLP, it is conceivable that, shortly after its isolation, the CLP may express *Pax5* or an equivalent commitment factor of the NK- and T-cell lineages, which subsequently prevents differentiation along myeloid lineages. In this view, the full developmental capacity of the CLP is only revealed under conditions that interfere with commitment to the lymphoid lineages, such as IL-7-dependent proliferation in the absence of *Pax5* function.

Multilineage priming of gene expression is a characteristic property of early haematopoietic progenitor cells⁴⁵. This process ensures that genes of different haematopoietic lineages are co-expressed within a single cell, albeit often at a low level³¹. The lineage-promiscuous gene expression observed in the *Pax5*^{-/-} pro-B cell thus provides molecular evidence for the progenitor status of this cell. The expression of *M-CSF-R*, *GM-CSF-Rα* and *Epo-R* may also explain why the *Pax5*^{-/-} pro-B cell can respond to different

cytokines. For instance, expression of the M-CSF receptor renders the *Pax5*^{-/-} pro-B cell responsive to M-CSF, which is produced by the ST2 feeder cells used for pro-B-cell culture¹⁹. IL-7 is, however, dominant over M-CSF in lymphoid precursor cells³⁸. As a consequence, M-CSF can support monocytic differentiation and proliferation only at limiting IL-7 concentrations, which explains the observed morphology change of *Pax5*^{-/-} pro-B cells in IL-7-depleted medium.

Early events in lineage specification involve the control of proliferation, survival and commitment of lineage-restricted precursor cells. The loss of an entire lineage by gene targeting can therefore result from interference with any one of these processes and does not necessarily imply a role of the mutated gene in lineage commitment. A hallmark of commitment is the stabilization of a lineage-specific gene-expression programme which permanently inhibits alternative lineage choices⁴⁵. Consequently, restriction of developmental plasticity provides a better criterion for defining lineage commitment. Loss-of-function experiments can therefore implicate a transcription factor in lineage commitment only if the lack of 'forward' differentiation is accompanied by maintenance of developmental plasticity in the transcription-factor-deficient progenitor cell. However, it must be possible to grow a mutant progenitor *in vitro* before its developmental potential can be analysed. So far, *Pax5* is the only lineage-specific transcription factor to fulfil this criterion, as *Pax5*^{-/-} pro-B cells can differentiate along multiple myeloid and lymphoid lineages with the exception of the B-cell pathway. Restoration of *Pax5* activity overcomes the B-cell developmental block, thus identifying *Pax5* as the critical B-lineage commitment factor. Surprisingly, considerable progression down the B-cell pathway is possible in the absence of commitment, as *Pax5*^{-/-} pro-B cells express many genes previously considered to be indicative of B-lineage commitment^{15,16}. These genes include *Igα(mb-1)*, *Igβ(B29)*, *VpreB* and *λ5* as well as the sterile transcripts of the *IgH* locus, all of which are under single or combinatorial control by the transcription factors E2A and EBF^{7-9,46}. Consistently, both E2A and EBF, which act upstream of *Pax5* in the genetic hierarchy of B-cell development^{7,8}, are equally expressed in *Pax5*^{-/-} and wild-type pro-B cells¹⁵. Hence, the *Pax5* mutation dissociates the initial activation of B-lineage-specific gene expression by E2A and EBF from the *Pax5*-dependent control of B-lineage commitment. *Pax5* fulfils a dual role in B-lineage commitment, as it activates the expression of B-cell-specific genes and simultaneously represses the lineage-promiscuous transcription of other haematopoietic genes. The transcriptional activation is best exemplified by *CD19*, whose expression is strictly dependent on *Pax5* (ref. 16) and should thus be regarded as a decisive marker of B-lineage commitment. Moreover, the *Pax5*-dependent repression of the *M-CSF-R* gene illustrates at the molecular level how the developmental potential is restricted at commitment by rendering B-cell precursors unresponsive to lineage-inappropriate cytokines such as M-CSF.

A long-standing debate concerns whether lineage commitment in the haematopoietic system occurs in a cell-autonomous, stochastic manner or is controlled by instructive signals from the local environment^{47,48}. Much attention has been paid to the role of haematopoietic growth factors in this process, but no evidence for any defect in lineage commitment has been observed despite the analysis of many growth factor gene knock-outs⁴⁸. Instead, growth factors seem to perform a selective role by promoting the survival and/or proliferation of independently committed cells. We have shown that the expression of *Pax5* is randomly initiated from only one of its two alleles at the onset of B-lymphopoiesis⁴⁹. The transcription of a single allele indicates that the lineage commitment gene *Pax5* is activated in a relatively inefficient, stochastic manner in the multipotent progenitor⁴⁹. However, once this stochastic event has occurred, it restricts the various developmental options of the progenitor to the B-cell pathway. Hence, our data

support the notion that B-lineage commitment is a stochastic rather than a deterministic process. □

Methods

Mice

Pax5^{-/-}, *RAG2*^{-/-} and *c-fos*^{-/-} mice^{14,17,29} were maintained on the C57BL/6x129/Sv background and genotyped as described^{5,15,29}.

Pro-B cell culture

Iscoe's modified Dulbecco's medium supplemented with 50 μM β-mercaptoethanol, 1 mM glutamine, 2% heat-inactivated fetal calf serum and 0.03% (w/v) primatone was used for all cell-culture experiments (referred to as IMDM medium). Pro-B cells isolated from bone marrow of 2-week-old mice were cultured on γ-irradiated ST2 cells in IMDM medium containing IL-7 (ref. 15).

Retroviral infection

The retrovirus pBabe-BSAP has been described¹⁶; pBabe-TRANCE was provided by K. Matsuo and E. F. Wagner. Murine *bcl2* cDNA was inserted into the histidinol resistance vector pMV-10 to generate a *bcl2*-expressing retrovirus. Transfected GP + E-86 packaging cells were selected with 2.5 μg ml⁻¹ puromycin or 10 mM histidinol (Sigma). Pro-B cells were infected by coculture with puromycin-histidinol-resistant ST2 cells and virus-producing packaging cells followed by puromycin (2.5 μg ml⁻¹) or histidinol (1 mM) selection.

Antibodies

Anti-CD40 (FGK45.5), anti-μH (M41), anti-CD19 (1D3), anti-c-Kit (ACK4) and anti-MHC class II (M5-114) antibodies were purified and biotinylated as described¹⁵. Fluorescein isothiocyanate (FITC)- and phycoerythrin (PE)-labelled anti-B220 (RA3-6B2), FITC-labelled anti-Ly49C (5E6) and anti-Mac-1 (M1/70), PE-conjugated CD11c (HL3) and Gr-1 (RB6-8C5), biotinylated anti-CD23 (B3B4) and anti-NK1.1 (PK136) antibodies and allophycocyanin (APC)-conjugated streptavidin were obtained from Pharmingen; PE-conjugated anti-F4/80 (C1A3-1) antibody from Serotec; and PE-labelled streptavidin from Southern Biotechnology Associates.

Flow cytometry

Cells were stained and analysed on a FACSCalibur (Becton Dickinson) as described¹⁴. Biotinylated antibodies were revealed with PE- or APC-streptavidin. Myeloid cells were incubated with 10% heat-inactivated rat serum before antibody staining.

Growth factors

Murine M-CSF (used at a final concentration of 25 ng ml⁻¹), G-CSF (25 ng ml⁻¹) and SCF (100 ng ml⁻¹) were purchased from R&D Systems. All other cytokines were produced by X63 or J558L cell transfectants⁵⁰ and used at a concentration of 1% (IL-7), 2% (IL-3, IL-6), 5% (IL-2, IL-4) and 10% (GM-CSF) conditioned supernatant.

Cell morphology assay

Individual B220⁺c-Kit⁺ pro-B cells from bone marrow (*ex vivo*) or long-term pro-B-cell cultures (*in vitro*) were sorted by a FACS Vantage TSO flow cytometer (Becton Dickinson) into single wells of 96-well plates containing ST2 cells in IL-7 medium. Colonies of small, round, refractile pro-B cells were identified after 10 days and thereafter scored for morphology changes defined by an increase in cell size, elongation and acquisition of granular vacuolated cytoplasm.

Cell-cloning assay

The cloning frequency of pro-B cells cultured on ST2 cells plus IL-7, in medium alone or in the presence of M-CSF, was assessed by limiting dilution analysis as described⁴. Cell numbers were normalized for viable cells (determined by trypan blue exclusion) before dilution into 96-well plates.

In vitro differentiation

All differentiation assays were performed with *in vitro* cultured pro-B cells. B lymphocytes: pro-B cells were incubated at 1–3 × 10⁶ cells per ml for three days in IMDM medium alone as described¹⁸. NK cells: pro-B cells were differentiated for 10 days in the presence of IL-2 and ST2 cells. Cytotoxicity assays were performed with YAC-1 cells or LPS-stimulated splenocytes as described²⁵. Macrophages: *Pax5*^{-/-} pro-B cells were differentiated in IMDM medium with ST2 cells for 10–14 days followed by terminal differentiation in M-CSF medium for seven days. Cells were transferred to chamber slides (Nalge Nunc), incubated with 10⁷ FITC-labelled heat-inactivated *E. coli* (Molecular Probes) for 1–2 h at 37 °C, washed twice in PBS, fixed in 2% paraformaldehyde for 10 min, air-dried, mounted in VectaShield (Vector Laboratories) and DABCO (10 μg μl⁻¹, Sigma)–DAPI (0.15 μg μl⁻¹, Sigma) solutions (1:1) and analysed on a fluorescence microscope (Zeiss Axiophot) with a CCD camera (Photometrics). Dendritic cells: *Pax5*^{-/-} pro-B cells were differentiated for 16 days on ST2 cells in GM-CSF medium. For immunocytochemical analysis, cells were flushed off the stromal cell layer and stained with PE-anti-MHC class II antibodies before cytospin centrifugation and fluorescence microscopy. Differentiated cells were tested for antigen presentation by the mixed leukocyte reaction as described²⁰. Osteoclasts: *Pax5*^{-/-}

pro-B cells were cultured for two weeks with M-CSF and ST2 cells expressing TRANCE. Cells were fixed in 3.7% formaldehyde for 30 min, dried and subjected to TRAP staining (Sigma). At day 10 of differentiation, cells were transferred onto osteologic bone discs (Millenium Biologic) together with ST2-TRANCE cells in M-CSF medium containing 100 nM dexamethasone, 10 mM vitamin D₃ and 50 µg ml⁻¹ vitamin C. Bone resorption was analysed after an additional 10-day incubation. Granulocytes: *Pax5*^{-/-} pro-B cells were cultured for three weeks in IMDM medium containing IL-3, IL-6 and SCF and then incubated for six days in IMDM medium containing G-CSF alone. The cell morphology was analysed by May-Grünwald-Giesma staining.

In vivo osteoclast development

Newborn mice were lethally irradiated (800 rad) and then intraperitoneally injected with a mixture of 10⁷ *in vitro* cultured *Pax5*^{-/-} pro-B cells and 10⁷ splenocytes from an adult *c-fos*^{-/-} mouse. Four weeks later, osteoclast formation was analysed by TRAP staining of isolated calvariae and histological sections of long bones³⁰.

RT-PCR

Total RNA (2 µg) from pro-B cells and lymphoid tissues was reverse-transcribed and PCR-amplified as described⁴⁹. The following primers were used: *Pax5-lacZ*, 5'-CATGGCGA-GAAGCTCTTTAGTTCC-3' and 5'-TGCAAGGCGATTAAGTTGGGTAAC-3', *CD3ε*, 5'-GTCTCCATCTCAGGAACAGT-3' and 5'-ATAGTCTGGGTTGGGAACAGG-3'. All other primers have been described^{15,25,28,31}. PCR products were identified on agarose gels by ethidium bromide staining or Southern hybridization.

Received 10 June; accepted 11 August 1999.

1. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
2. Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Kayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* **173**, 1213–1225 (1991).
3. Rolink, A., Grawunder, U., Winkler, T. H., Karasuyama, H. & Melchers, F. IL-2 receptor α chain (CD25,TAC) expression defines a crucial stage in pre-B cell development. *Int. Immunol.* **6**, 1257–1264 (1994).
4. Rolink, A., Kudo, A., Karasuyama, H., Kikuchi, Y. & Melchers, F. Long-term proliferating early pre B cell lines and clones with the potential to develop to surface Ig-positive, mitogen reactive B cells *in vitro* and *in vivo*. *EMBO J.* **10**, 327–336 (1991).
5. Thévenin, C., Nutt, S. L. & Busslinger, M. Early function of Pax5 (BSAP) prior to the pre-B cell receptor stage of B lymphopoiesis. *J. Exp. Med.* **188**, 735–744 (1998).
6. Zhuang, Y., Soriano, P. & Weintraub, H. The helix-loop-helix gene E2A is required for B cell formation. *Cell* **79**, 875–884 (1994).
7. Bain, G. *et al.* E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* **79**, 885–892 (1994).
8. Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**, 263–267 (1995).
9. Sigvardsson, M., O'Riordan, M. & Grosschedl, R. EBF and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes. *Immunity* **7**, 25–36 (1997).
10. Kee, B. L. & Murre, C. Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12. *J. Exp. Med.* **188**, 699–713 (1998).
11. Busslinger, M. & Nutt, S. L. in *Molecular Biology of B-Cell and T-Cell Development* (eds Monroe, J. G. & Rothenberg, E. V.) 83–110 (Humana, Totowa, New Jersey, 1998).
12. Adams, B. *et al.* Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev.* **6**, 1589–1607 (1992).
13. Li, Y.-S., Wasserman, R., Hayakawa, K. & Hardy, R. R. Identification of the earliest B lineage stage in mouse bone marrow. *Immunity* **5**, 527–535 (1996).
14. Urbánek, P., Wang, Z.-Q., Fetka, I., Wagner, E. F. & Busslinger, M. Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79**, 901–912 (1994).
15. Nutt, S. L., Urbánek, P., Rolink, A. & Busslinger, M. Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the *IgH* locus. *Genes Dev.* **11**, 476–491 (1997).
16. Nutt, S. L., Morrison, A. M., Dörfler, P., Rolink, A. & Busslinger, M. Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments. *EMBO J.* **17**, 2319–2333 (1998).
17. Shinkai, Y. *et al.* RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855–867 (1992).
18. Rolink, A., Melchers, F. & Andersson, J. The SCID but not the RAG-2 gene product is required for μ -Se heavy chain class switching. *Immunity* **5**, 319–330 (1996).
19. Yamane, T. *et al.* Development of osteoclasts from embryonic stem cells through a pathway that is *c-fms* but not *c-kit* dependent. *Blood* **90**, 3516–3523 (1997).
20. Inaba, K. *et al.* Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **176**, 1693–1702 (1992).

21. Branchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. *Nature* **392**, 245–252 (1998).
22. Kong, Y.-Y. *et al.* OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* **397**, 315–323 (1999).
23. Liu, F., Poursine-Laurent, J., Wu, H. Y. & Link, D. C. Interleukin-6 and the granulocyte colony-stimulating factor receptor are major independent regulators of granulopoiesis *in vivo* but are not required for lineage commitment or terminal differentiation. *Blood* **90**, 2583–2590 (1997).
24. Ogasawara, K. *et al.* Requirement for IRF-1 in the microenvironment supporting development of natural killer cells. *Nature* **391**, 700–703 (1998).
25. Rolink, A. *et al.* A subpopulation of B220⁺ cells in murine bone marrow does not express CD19 and contains natural killer cell progenitors. *J. Exp. Med.* **183**, 187–194 (1996).
26. Raulet, D. H. Development and tolerance of natural killer cells. *Curr. Opin. Immunol.* **11**, 129–134 (1999).
27. Liao, N.-S., Bix, M., Zijlstra, M., Jaenisch, R. & Raulet, D. MHC class I deficiency: susceptibility to natural killer (NK) cells and impaired NK activity. *Science* **253**, 199–202 (1991).
28. Rolink, A. G., Nutt, S. L., Melchers, F. & Busslinger, M. Long-term *in vivo* reconstitution of T cell development by Pax5-deficient B cell progenitors. *Nature* **401**, 603–606 (1999).
29. Wang, Z.-Q. *et al.* Bone and haematopoietic defects in mice lacking *c-fos*. *Nature* **360**, 741–745 (1992).
30. Grigoriadis, A. *et al.* c-Fos: a key regulator of osteoclast-macrophage lineage determination and bone remodeling. *Science* **266**, 443–448 (1994).
31. Hu, M. *et al.* Multilineage gene expression precedes commitment in the hematopoietic system. *Genes Dev.* **11**, 774–785 (1997).
32. Borrello, M. A. & Phipps, R. P. The B/macrophage cell: an elusive link between CD5⁺ B lymphocytes and macrophages. *Immunol. Today* **17**, 471–475 (1996).
33. Akashi, K. *et al.* Simultaneous occurrence of myelomonocytic leukemia and multiple myeloma: involvement of common leukemic progenitors and their developmental abnormality of “lineage infidelity”. *J. Cell Physiol.* **148**, 446–456 (1991).
34. Klinken, S. P., Alexander, W. S. & Adams, J. M. Hematopoietic lineage switch: v-*raf* oncogene converts Eμ-*myc* transgenic B cells into macrophages. *Cell* **53**, 857–867 (1988).
35. Strasser, A., Elefanti, A. G., Harris, A. W. & Cory, S. Progenitor tumours from Eμ-*bcl-2*-*myc* transgenic mice have lymphomeloid differentiation potential and reveal developmental differences in cells survival. *EMBO J.* **15**, 3823–3824 (1996).
36. Davidson, W. F., Pierce, J. H., Rudikoff, S. & Morse III, H. C. Relationships between B cell and myeloid differentiation: studies with a B lymphocyte progenitor line, HAFTL-1. *J. Exp. Med.* **168**, 389–407 (1988).
37. Principato, M. *et al.* Transformation of murine bone marrow cells with combined v-*raf*-v-*myc* oncogenes yields clonally related mature B cells and macrophages. *Mol. Cell. Biol.* **10**, 3562–3568 (1990).
38. Borzillo, G. V., Ashmun, R. A. & Sherr, C. J. Macrophage lineage switching of murine early pre-B lymphoid cells expressing transduced *fms* genes. *Mol. Cell. Biol.* **10**, 2703–2714 (1990).
39. Cumano, A., Paige, C. J., Iscove, N. N. & Brady, G. Bipotential precursors of B cells and macrophages in murine fetal liver. *Nature* **356**, 612–615 (1992).
40. Kawamoto, H., Ohmura, K. & Katsura, Y. Direct evidence for the commitment of hematopoietic stem cells to T, B and myeloid lineages in murine fetal liver. *Int. Immunol.* **9**, 1011–1019 (1997).
41. Aiba, Y. & Ogawa, M. Development of natural killer cells, B lymphocytes, macrophages, and mast cells from single hematopoietic progenitors in culture of murine fetal liver cells. *Blood* **90**, 3923–3930 (1997).
42. Lacaud, G., Carlsson, L. & Keller, G. Identification of a fetal hematopoietic precursor with B cell, T cell, and macrophage potential. *Immunity* **9**, 827–838 (1998).
43. Galy, A., Travis, M., Cen, D. & Chen, B. Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. *Immunity* **3**, 459–473 (1995).
44. van Freeden-Jeffry, U. *et al.* Lymphopenia in interleukin (IL)-7 gene-deleted mice identifies IL-7 as a nonredundant cytokine. *J. Exp. Med.* **181**, 1519–1526 (1995).
45. Enver, T. & Greaves, M. Loops, lineages, and leukemia. *Cell* **94**, 9–12 (1998).
46. Åkerblad, P., Rosberg, M., Leanderson, T. & Sigvardsson, M. The *B29* (immunoglobulin β-chain) gene is a genetic target for early B-cell factor. *Mol. Cell. Biol.* **19**, 392–401 (1999).
47. Metcalf, D. Lineage commitment and maturation in hematopoietic cells: the case for extrinsic regulation. *Blood* **92**, 345–348 (1998).
48. Enver, T., Heyworth, C. M. & Dexter, T. M. Do stem cells play dice? *Blood* **92**, 348–351 (1998).
49. Nutt, S. L. *et al.* Independent regulation of the two Pax5 alleles during B-cell development. *Nature Genet.* **21**, 390–395 (1999).
50. Karasuyama, H. & Melchers, F. Establishment of mouse cell lines which constitutively secrete large quantities of interleukin 2, 3, 4 and 5, using modified cDNA expression vectors. *Eur. J. Immunol.* **18**, 97–104 (1988).

Acknowledgements

We thank K. Matsuo and E. F. Wagner for *c-fos*^{-/-} mice, ST2-TRANCE cells and advice concerning osteoclast biology; F. Alt for *RAG2*^{-/-} mice; P. Steinlein and M. Dessing for FACS sorting and assistance with photography; and H. Beug for critical reading of the manuscript. This work was supported by the I.M.P. Vienna, by a grant from the Austrian Industrial Research Promotion Fund and by the Basel Institute for Immunology.

Correspondence and requests for material should be addressed to M.B. (e-mail: busslinger@nt.imp.univie.ac.at).

Low-temperature crystallization of silicate dust in circumstellar disks

F. J. Molster*, I. Yamamura*, L. B. F. M. Waters*†, A. G. G. M. Tielens‡, Th. de Graauw‡, T. de Jong*§, A. de Koter*, K. Malfait†, M. E. van den Ancker*, H. van Winckel†, R. H. M. Voors§ & C. Waelkens†

* Astronomical Institute 'Anton Pannekoek', University of Amsterdam, Kruislaan 403, NL-1098 SJ Amsterdam, The Netherlands

† Instituut voor Sterrenkunde, KU Leuven, Celestijnenlaan 200B, 3001 Heverlee, Belgium

‡ SRON Laboratory for Space Research Groningen, PO Box 800, NL-9700 AV Groningen, The Netherlands

§ SRON Laboratory for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands

|| Astronomical Institute, University of Utrecht, Princetonplein 5, 3508 TA Utrecht, The Netherlands

Silicate dust in the interstellar medium is observed to be amorphous¹, yet silicate dust in comets^{2,3} and interplanetary dust particles⁴ is sometimes partially crystalline. The dust in disks that are thought to be forming planets around some young stars^{5,6} also appears to be partially crystalline. These observations suggest that as the dust goes from the precursor clouds to a planetary system, it must undergo some processing, but the nature and extent of this processing remain unknown. Here we report observations of highly crystalline silicate dust in the disks surrounding binary red-giant stars. The dust was created in amorphous form in the outer atmospheres of the red giants, and therefore must be processed in the disks to become crystalline. The temperatures in these disks are too low for the grains to anneal; therefore, some low-temperature process must be responsible. As the physical properties of the disks around young stars and red giants are similar, our results suggest that low-temperature crystallization of silicate grains also can occur in protoplanetary systems.

The 'life cycle' of cosmic dust commences when solid particles condense in the outflows of evolved red giants. Most of the dust grains that form in these outflows have a silicate composition whose basic building block is an SiO₄ tetrahedron with cations of iron, magnesium and other metals. Observations show that these grains have a disordered amorphous lattice structure, with only a minor fraction in an ordered (crystalline) form^{7,8}. The grains subsequently enter the interstellar medium, and finally become the building blocks of a new generation of stars and planetary systems. Observations of dust in the interstellar medium and in regions of active star formation show that the structure of silicates in the interstellar medium remains essentially amorphous¹.

Silicates with a highly ordered, crystalline lattice structure are known to exist in Solar System comets^{2,3} and in interplanetary dust particles⁴, and should reflect the composition of dust in the primitive protosolar nebula. In addition, crystalline silicates have been found in the dusty disks surrounding young stars^{5,6}. These disks also contain a population of large dust grains whose size significantly exceeds that of dust in the interstellar medium. These large grains are formed in the disk through coagulation of smaller particles. The crystalline dust must have formed in the protoplanetary disk, but the physical process responsible for its formation is not known. Annealing by means of heating to temperatures above the glass temperature (~1,000 K) requires the dust grains to have been close to the star. Their inclusion in icy comets, however, implies significant radial mixing of dust in the disk, as comets are believed to form in the cool outer regions of the disk. Models for the evolution of the protosolar cloud indeed suggest that radial mixing may be important^{9,10}. Here we present observations of crystalline

dust in disks surrounding red giants and related objects, demonstrating that crystallization at low temperatures also occurs. Such crystallization allows for the inclusion of crystalline silicates in comets and interplanetary dust particles without the need for radial mixing.

We have investigated the composition of silicate dust in the surroundings of red giants in binary systems and similar evolved stars using 2–45 μm spectra obtained with the Short Wavelength Spectrometer (SWS)¹¹ on board the Infrared Space Observatory (ISO)¹². The spectrum of the carbon-rich giant IRAS09425–6040 (Fig. 1) shows by far the largest percentage of crystalline silicates (75%) observed to date. These silicates are the processed remnants of a previous mass-loss phase when the star was still oxygen-rich. The spectral features of dust surrounding the binary AC Herculis are very similar to those seen in Solar System comets^{2,3}, such as Hale-Bopp (Fig. 1), and to those of some protoplanetary disks surrounding young stars^{5,6}. On the basis of the compositional similarities of the dust in disks surrounding old and young stars, we conclude that the processes which are responsible for the crystallization of silicates are likely to be the same.

Several objects in the ISO sample with a high degree of crystallinity show unusually high millimetre-wavelength fluxes (F_{mm}) compared to their IRAS 60- μm fluxes (F_{60} ; Fig. 2). Cool giants and well-known red giants such as OH26.5+0.6 are characterized by large F_{60}/F_{mm} ratios. A low F_{60}/F_{mm} ratio indicates the presence of large, cold grains, as small grains (<10 μm in size) radiate inefficiently at wavelengths that are large compared to their size. It is likely that these large grains have been formed by coagulation

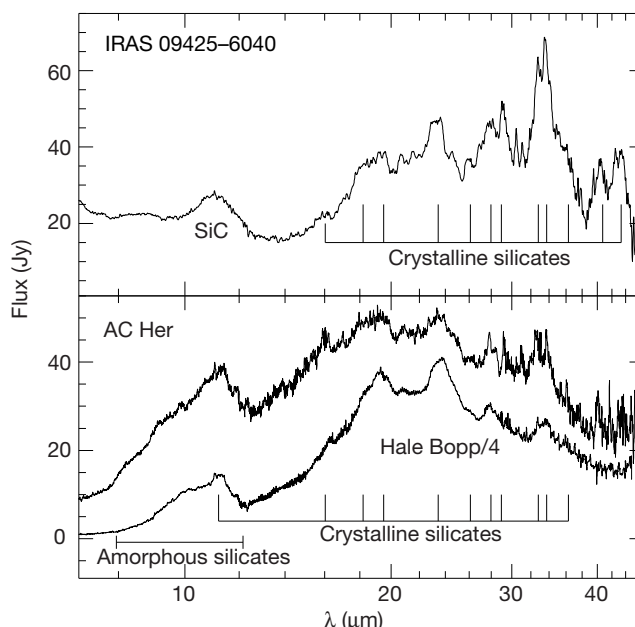


Figure 1 Infrared spectra showing crystalline silicate features. ISO-SWS grating scan of the carbon-rich giant star IRAS09425–6040 (top panel), taken on 27 July 1996, and of the post-AGB star AC Her (bottom panel), taken on 18 April 1997. Note the prominent crystalline silicate bands. The ISO-SWS spectrum of comet Hale–Bopp³ is shown for comparison. At short wavelengths ($\lambda < 14 \mu\text{m}$), the spectrum of IRAS09425–6040 is characterized by carbonaceous compounds: molecular bands of C₂H₂, HCN and C₃ (not shown; F.J.M. *et al.*, manuscript in preparation), and the SiC dust feature at 11 μm , all originating in the present-day carbon-rich photosphere and outflow from this object. At long wavelengths ($\lambda > 14 \mu\text{m}$) the spectrum is completely dominated by emission from cool, highly crystalline oxygen-rich dust, 75% of which is forsterite and enstatite (F.J.M. *et al.*, manuscript in preparation). AC Her is a binary, of which one component has recently left the red-giant phase and is now rapidly evolving to higher temperatures²⁸. The silicate features of AC Her are a blend of amorphous and crystalline silicates (~25%), and are very similar to those seen in comet Hale–Bopp.

in a long-lived circumstellar disk. Even in the modest ISO sample, we already find several examples of disks around evolved stars, as evidenced by direct imaging or from the shape of the spectral energy distribution. The object AC Her is a member of a class of evolved giants (some and possibly all of which are binaries) with similar circumstellar gas and dust properties. We conclude that the presence of disks around evolved binary stars, and the formation of planet-like bodies in these disks, may be widespread. The sources in our sample with a high abundance of crystalline silicates all have a population of large, cold grains, and show evidence for a disk. However, coagulation does not necessarily imply a high degree of crystallization, as some sources show a low F_{60}/F_{mm} ratio without prominent crystalline silicate emission¹³.

The high degree of crystallization of the oxygen-rich material in disks surrounding red giants is remarkable. The question arises as to what controls the efficiency of crystallization in the disk. It is very unlikely that the dust composition seen today is representative of the dust composition when the dust grains formed in the oxygen-rich red-giant wind, as red giants that are observed produce predominantly amorphous silicates. All sources with a high abundance of crystalline silicates do show evidence for a disk (Fig. 2). We therefore attribute the high degree of crystallinity to processing

during the long storage time in the stationary disk. Two ways of producing highly crystalline dust disks can be considered: selective removal of the small amorphous grains, and crystallization of the amorphous grains.

The small amorphous grains may have been selectively removed by processes similar to those occurring in our Solar System (radiation pressure including Poynting–Robertson drag in a gas-poor dust disk). However, this would require an unrealistically high initial mass of the disk (for instance, in the case of IRAS09425–6040, we estimate an original total disk mass of 0.1 solar masses, assuming an initially normal ratio of amorphous to crystalline silicate; F.J.M. *et al.*, manuscript in preparation). Alternatively, most (small) amorphous grains may not have been physically removed, but rather coagulated into the large grains which are now so evident in the millimetre flux. The physical cause of such a preferential coagulation of amorphous grains is at present unclear. In the Solar System, inclusions of crystalline silicates in interplanetary dust particles^{4,14} demonstrate that these grains can efficiently take part in the coagulation process of protoplanetary disks.

Evidence against high temperatures as the cause of crystallization of amorphous silicates in stationary disks can be found from ISO-SWS observations of the carbon-rich star V778 Cygni¹⁵. This object may have a stationary circumstellar disk of hot amorphous silicates at a temperature of 600 K. No indication of the presence of crystalline silicates was found in the spectrum. Indeed, while circumstellar amorphous silicates show a range of temperatures up to 1,000 K, a low dust temperature (100–250 K) is a general characteristic of all crystalline silicates observed to date. Therefore the observational evidence suggests that the crystallization process is not driven by the thermal (radiation) temperature of the dust, even at relatively high temperatures. This conclusion is supported by a comparison of the age of the dust disks in evolved stars with the annealing timescales of amorphous grains, assuming that thermal processes dominate the crystallization: while the age of the disks is estimated to be $\sim 10^5$ years, the annealing timescales predicted from kinetic theory at temperatures of 100–250 K are many orders of magnitude larger^{16,17}. Nevertheless, longevity seems to be a prerequisite for crystallization.

Although the origin of the disks around evolved stars (mass loss) and young stars (accretion from the interstellar matter) are different, the similarity in their properties (such as coagulation, gas depletion (ref. 18, and F.J.M. *et al.*, manuscript in preparation) and the presence of crystalline silicates in a long-lived debris disk) are remarkable. Radial mixing is not expected to occur in the disks around evolved stars, as these are not accretion disks; however, our observations indicate that athermal crystallization (defined here as crystallization at temperatures below the glass temperature) may be responsible for the presence of crystalline silicates in planet-forming disks, without the need for radial mixing.

We now consider the physical and chemical processes responsible for the crystallization at low temperatures. Possibly, crystallization is initiated by an athermal process, and the crystallization front is driven either by the release of the latent heat of crystallization through the whole grain, or by mechanical energy introduced by this energetic event. Both processes are similar to what is seen for silicon^{19,20}. Alternatively, a local, temporal heating process (such as lightning) may raise the temperature sufficiently to cause crystallization, followed by slow (~ 1 s) cooling in a highly optically thick and turbulent disk environment: a process which may have led to the production of chondrules in the solar nebula²¹ (but here for much smaller grains). Grain–grain collisions at relative velocities less than 20 km s^{-1} will drive a strong shock front in the projectile and target which could cause local or complete melting²². In principle, this could lead to the formation of (new) crystal structures, as happens for example in carbon-based materials where graphite is transformed into diamond by shock loading. But for silicate materials, experiments show that on cooling the material solidifies in an amorphous form²².

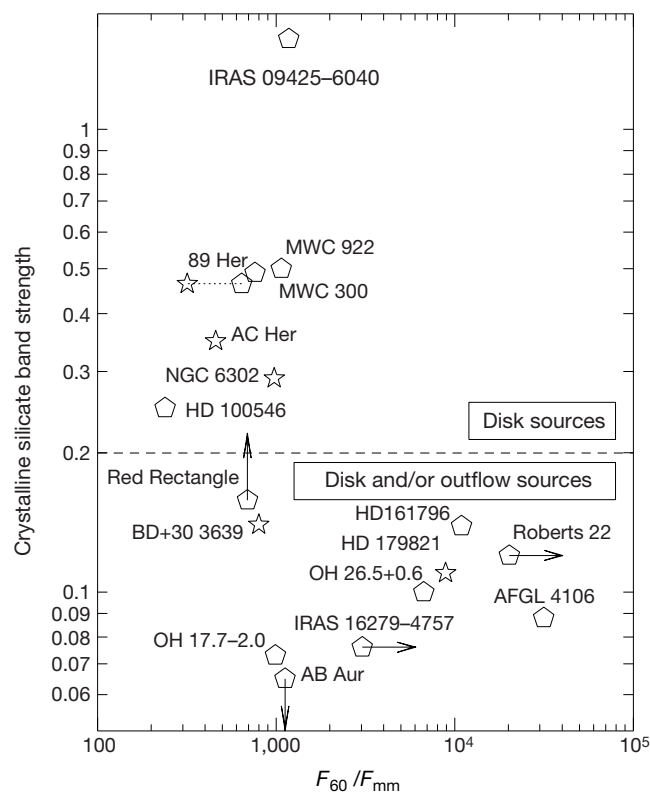


Figure 2 Correlation between the ratio of the IRAS 60- μm flux to the millimetre-wave flux and the crystalline silicate band strength. F_{mm} is measured at 1.1 mm wavelength (star data points) or 1.3 mm wavelength (pentagons). The crystalline band strength is measured from the crystalline forsterite band at $33.6 \mu\text{m}$ ($(F_{33.6 \mu\text{m peak}} - F_{\text{cont}})/F_{\text{cont}}$). The selected sources have dust colour temperatures above ~ 100 K, so that the Planck function peaks at wavelengths $< 60 \mu\text{m}$. The emptiness of the upper right corner of this diagram is an indication that the presence of highly crystalline dust is related to disks and grain-growth. All stars above the dashed line ($(F_{33.6 \mu\text{m peak}} - F_{\text{cont}})/F_{\text{cont}} = 0.2$) have relatively small F_{60}/F_{mm} ratios, and have indications of the presence of a disk (from, for example, direct imaging and/or the spectral energy distribution). The stars below this line are predominantly normal outflow sources, without any evidence for a disk, although exceptions exist (such as Roberts 22). NGC6302 and BD+30 3639 have been shifted by a factor of two along the x-axis as only about half of the millimetre-wave continuum flux is due to dust emission while the other half is due to free–free emission.

Energetic ion and electron beams can induce recrystallization of amorphous layers in semiconductors at temperatures well below those that are required for thermal annealing^{20,23}. There are few laboratory studies on this process for the astrophysically more relevant silicate grains. However, observations of the effect of solar flares on the structure of interplanetary dust particles indeed show evidence for recrystallization rims, indicating that these energetic particles can trigger local crystallization of the lattice²⁴. These observations underpin the importance of crystallization at low temperatures as an alternative to thermal annealing followed by radial mixing as a mechanism for the formation of crystalline silicates in circumstellar disks.

A final consideration is the chemical composition of the crystalline silicates. In all our studies, we find that these are very Mg-rich with less than 1% of Fe, Ni and Ca (refs 5, 8, and F.J.M. *et al.*, and T. Lim *et al.*, manuscripts in preparation). To explain the absorption properties of amorphous silicates, it is often assumed that they have a significant abundance of metals such as Fe, Ni and Ca (ref. 25). The difference in absorption properties is evidenced by the differences in temperature for the amorphous and crystalline silicates in, for example, NGC6302 (T. Lim *et al.*, manuscript in preparation) and AFGL4106 (ref. 26). This implies that these metals are removed during the crystallization process. Although no laboratory studies have been done for silicate grains, the process of 'impurity' removal together with low-temperature crystallization has been found in laboratory studies of semiconductors²⁷. Any physical mechanism for crystallization must be able to account for these differences in chemical composition between the amorphous and crystalline silicates. It would be useful to make laboratory measurements of the effects of irradiation and local heating on the crystallization process of silicates. □

Received 7 June; accepted 28 July 1999.

- Dorschner, J. & Henning, T. Dust metamorphosis in the galaxy. *Astron. Astrophys. Rev.* **6**, 271–333 (1995).
- Hanner, M. S., Lynch, D. K. & Russel, R. W. The 8–13 micron spectra of comets and the composition of silicate grains. *Astrophys. J.* **425**, 274–285 (1994).
- Crovisier, J. *et al.* The spectrum of comet Hale–Bopp (C/1995 01) observed with the Infrared Space Observatory at 2.9 astronomical units from the Sun. *Science* **275**, 1904–1907 (1997).
- Bradley, J. P., Snow, T. P., Brownlee, D. E. & Hanner, M. S. In *Solid Interstellar Matter: The ISO Revolution* (eds d'Hendecourt, L., Joblin, C. & Jones, A.) 297–315 (Centre de Physique des Houches No. 11, Springer, Berlin, 1999).
- Waelkens, C., Malfait, K. & Waters, L. B. F. M. ISO observations of isolated Herbig Ae/Be stars. *Astrophys. Space Sci.* **255**, 25–33 (1998).
- Malfait, K. *et al.* The spectrum of the young star HD 100546 observed with the Infrared Space Observatory. *Astron. Astrophys.* **332**, L25–L28 (1998).
- Waters, L. B. F. M. *et al.* Mineralogy of oxygen-rich dust shells. *Astron. Astrophys.* **315**, L361–L364 (1996).
- Waters, L. B. F. M. & Molster, F. J. Crystalline silicates in AGB and post-AGB stars. In *Asymptotic Giant Branch Stars* (eds Le Bertre, T., Lèbre, A. & Waelkens, C.) 209–219 (IAU Symp. 191, Astronomical Society of the Pacific, San Francisco, 1999).
- Shu, F. H., Shang, H. & Lee, T. Toward an astrophysical theory of chondrites. *Science* **271**, 1545–1552 (1996).
- Shu, F. H., Shang, H., Glassgold, A. E. & Lee, T. X-rays and fluctuating X-winds from protostars. *Science* **277**, 1475–1479 (1997).
- De Graauw, Th. *et al.* Observing with the ISO Short-Wavelength Spectrometer. *Astron. Astrophys.* **315**, L49–L54 (1996).
- Kessler, M. F. *et al.* The Infrared Space Observatory (ISO) mission. *Astron. Astrophys.* **315**, L27–L31 (1996).
- van den Ancker, M. E. *et al.* ISO spectroscopy of circumstellar dust in the Herbig Ae systems AB Aur and HD 163296. *Astron. Astrophys.* (in the press).
- Bradley, J. P. Chemically anomalous preaccretionally irradiated grains in interplanetary dust from comets. *Science* **265**, 925–929 (1994).
- Yamamura, I. *et al.* In *Proc. 1st ISO Workshop on Analytical Spectroscopy* (eds Heras, A. M. *et al.*) 313–314 (SP-419, ESA, Noordwijk, 1997).
- Hallenbeck, S. & Nuth, J. Infrared observations of the transition from chaotic to crystalline silicates via thermal annealing in the laboratory. *Astrophys. Space Sci.* **255**, 427–433 (1998).
- Gail, H.-P. Chemical reactions in protoplanetary accretion disks (IV. Multicomponent dust mixture). *Astron. Astrophys.* **322**, 1099–1122 (1998).
- Jura, M., Balm, S. P. & Kahane, C. A long-lived disk around the Red Rectangle? *Astrophys. J.* **453**, 721–726 (1995).
- Galvin, G. J. *et al.* Measurement of the velocity of the crystal-liquid interface in pulsed laser annealing of Si. *Phys. Rev. Lett.* **48**, 33–36 (1981).
- Grun, J. *et al.* Athermal annealing of silicon. *Phys. Rev. Lett.* **78**, 1584–1587 (1997).
- Wood, J. A. In *Protostars & Planets II* (eds Clark, D. C. & Matthews, M. S.) 687–702 (Univ. Arizona Press, Tucson, 1985).

- Tielens, A. G. G. M., McKee, C. F., Seab, C. G. & Hollenbach, D. J. The physics of grain-grain collisions and gas-grain sputtering in interstellar shocks. *Astrophys. J.* **431**, 321–340 (1994).
- Williams, J. S., Elliman, R. G., Brown, W. L. & Seidel, T. E. Dominant influence of beam-induced interface rearrangements on solid-phase epitaxial crystallization of amorphous silicon. *Phys. Rev. Lett.* **55**, 1482–1485 (1985).
- Thiel, K., Bradley, J. & Sphor, R. Investigations of solar flare tracks in IDP's: Some recent results. *Int. J. Radiat. Appl. Instrum. D* **19**, 709–716 (1991).
- Jones, T. W. & Merrill, K. M. Model dust envelopes around late-type stars. *Astrophys. J.* **209**, 509–524 (1976).
- Molster, F. J. *et al.* The composition and nature of the dust shell surrounding the binary AFG4106. *Astron. Astrophys.* (in the press).
- Elliman, R. G., Williams, J. S., Maher, D. M. & Brown, W. L. Kinetics, microstructure and mechanisms of ion beam induced epitaxial crystallization of semiconductors. *Mater. Res. Soc. Symp. Proc.* **51**, 319–327 (1986).
- Van Winckel, H. *et al.* The binary RV Tauri star AC Her and evidence for a long-lived dust-disc. *Astron. Astrophys.* **336**, L17–L20 (1998).

Acknowledgements

We thank T. Henning and his group for discussions about crystals and crystallization, and R. D. Schuiling for sharing his geophysical knowledge of the crystallization of silicates. F.J.M., I.Y., L.B.F.M.W., A.d.K. and M.E.v.d.A. acknowledge support from NWO grants; C.W. acknowledges support from the Belgian Federal Services for Scientific, Technological and Cultural Affairs, and from the Onderzoeksfonds KU Leuven. H.v.W. and K.M. acknowledge support from the Fund for Scientific Research of Flanders. This work was partly supported by a NWO Spinoza grant to E.P.J. van den Heuvel. This work is based partly on observations with ISO, an ESA project with instruments funded by ESA Member States (especially the PI countries: France, Germany, The Netherlands and the United Kingdom) and with the participation of ISAS and NASA. It also involves observations obtained at the European Southern Observatory, La Silla, Chile.

Correspondence and requests for materials should be addressed to F.J.M. (e-mail: frankm@astro.uva.nl).

Discovery of a moon orbiting the asteroid 45 Eugenia

W. J. Merline*, L. M. Close†, C. Dumas‡, C. R. Chapman*, F. Roddier§, F. Ménard||, D. C. Slater¶, G. Duvert#, C. Shelton* & T. Morgan**

* Southwest Research Institute, Boulder, Colorado 80302, USA

† European Southern Observatory, D-85748 Garching, Germany

‡ Jet Propulsion Laboratory, California Institute of Technology, MS 183-501, Pasadena, California 91109, USA

§ Institute for Astronomy, University of Hawaii, Honolulu, Hawaii 96822, USA

|| Canada-France-Hawaii Telescope Corporation, Kamuela, Hawaii 96743, USA

¶ Southwest Research Institute, San Antonio, Texas 78228, USA

Laboratoire d'Astrophysique de Grenoble, BP53, F-38041 Grenoble Cedex 9, France

* W. M. Keck Observatory, Kamuela, Hawaii 96743, USA

** NASA Headquarters, Washington DC 20546, USA

Evidence for asteroidal satellites (moons) has been sought for decades, because the relative frequency of such satellites will bear on the collisional history of the asteroid belt and the Solar System, yet only one has been detected unambiguously^{1–3}. Here we report the discovery of a satellite of the asteroid 45 Eugenia, using an adaptive optics system on a ground-based telescope. The satellite has a diameter of about 13 km, and an orbital period of about 4.7 days with a separation of 1,190 km from Eugenia. Using a previously determined⁴ diameter for Eugenia, we estimate that its bulk density is about 1.2 g cm^{−3}, which is similar to that of the C-type asteroid Mathilde^{5,6}. This implies that Eugenia, also a low-albedo C-type asteroid, may be a rubble pile, or composed of primitive, icy materials of low bulk density.

The possibility that some asteroids have satellites has been debated since the 1970s (ref. 7), when observers reported anomalous, generally unconfirmed, 'additional' events during occultations of bright stars by asteroids, sometimes interpreted to be caused by

unseen companions. Asteroidal satellites have been suspected from unusual light curves, slow rotation rates, and double impact craters on planetary surfaces. The prevalence of asteroidal satellites, inferred observationally and theoretically, has ranged from common⁸, to uncommon⁷, to essentially absent⁹. Stable orbits for satellites can exist, even in the presence of solar or jovian perturbations¹⁰. The first asteroidal satellite to be found was Ida's Dactyl, discovered during the Galileo spacecraft's 1993 fly-by¹⁻³. Subsequent modelling of the catastrophic collisional formation, tidal evolution, and longevity of asteroidal satellites suggests that such satellites may be common^{11,12}.

Study of asteroidal satellites (or their absence) will reveal much about the processes that formed the asteroids and collections of asteroids, known as 'families', and their collisional history, which, in turn, helps us understand the formation and collisional history of the planetary system^{13,14}. We will also learn about the physical structures and compositions of the primary asteroids because the presence of satellites permits determination of their masses and bulk densities (given knowledge of the asteroid's size).

Direct imaging of asteroidal satellites has been hampered by lack of adequate angular resolution to distinguish objects separated by

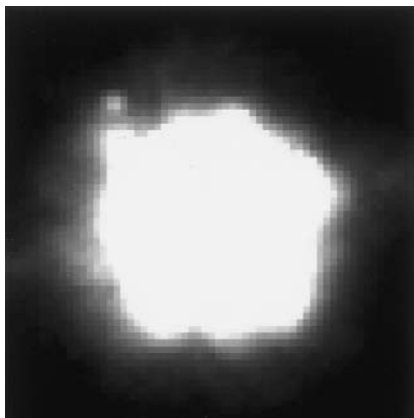


Figure 1 Discovery image, showing the asteroid 45 Eugenia and its satellite S/1998(45)1. This is a non-deconvolved summation of 16 raw images of 15-s duration each, obtained on 1 November 1998 UT. The observations were made with the 3.6-m Canada-France-Hawaii Telescope (CFHT), using the 'PUEO' adaptive optics (AO) bonette (adaptor). AO compensates atmospheric image degradation using a fast (kHz rate) deformable mirror. Unlike early AO systems developed for defence applications, PUEO is based on the concept of wave-front-curvature sensing and compensation³³ and is one of a new class of AO systems, first developed at the University of Hawaii³⁴. This 19-actuator AO system requires a natural 'guide' source (star or asteroid), as faint as $V = 17$ mag, to sense the wave-front aberrations. Best results are obtained with a guide source brighter than $V = 13$ mag, in which case PUEO can deliver diffraction-limited stellar images at near-infrared wavelengths. At a wavelength of $1.65 \mu\text{m}$, the image full-width at half-maximum (FWHM) is typically $0.12''$ and has a Strehl ratio (the ratio of peak brightness of acquired image to the peak brightness of a perfectly diffraction-limited point-source) of ~ 0.5 (ref. 35). Images were recorded with the CFHT near-infrared camera—a $1,024 \times 1,024$ HgCdTe focal plane array, sensitive over $0.7\text{--}2.5 \mu\text{m}$. The plate scale is 0.0350 ± 0.0003 arcsec per pixel, yielding a total field-of-view of $36'' \times 36''$. Each image set consists of four different dither positions (in a 2×2 14-arcsec offset pattern) on the detector. In each dither position, four 15-s images were obtained, resulting in final on-source integration time of 4 min for each final image. Each sky-subtracted, flat-fielded frame was co-aligned to within 0.01 pixels by cross-correlation of the asteroid. The 16 frames were then averaged, weighted by the peak flux in the asteroid core to produce a final image, so that frames with the least atmospheric attenuation and best AO compensation were given highest weight. In total we obtained 16 H-band ($1.65 \mu\text{m}$) averaged-images of 45 Eugenia ($V \approx 12.4$ mag) over a 10-day period (1–10 November 1998 UT; 12 of these were of high quality (AO-compensated seeing of $0.12\text{--}0.15$ FWHM)).

fractions of an arcsecond and by lack of sufficient dynamic range of detectors to resolve brightness differences of many magnitudes. Our work demonstrates the tremendous power of a new technology, adaptive optics, to reduce blurring caused by the Earth's atmosphere, enabling diffraction-limited imaging using large ground-based telescopes, and opening new regimes of Solar System and astronomical science.

In 1998, we began to survey 200 asteroids for companions. Our strategy was to study those asteroids that were targets of spacecraft missions, those previously suspected of having satellites, and those that are both bright and close enough to Earth that their gravitational sphere of influence subtends a significant angle. For these observations, we used the Canada-France-Hawaii Telescope.

The satellite was discovered during our first observing night, as a faint point-source within $0.76''$ of Eugenia (Fig. 1). We then switched Eugenia to a higher-than-survey priority to confirm the detection and determine orbital characteristics. On both 7 and 10 November 1998, the satellite was detected during four separate sessions. The location of each detection is shown in Fig. 2, indicating both that we can track motion within a night and that there can be no ambiguity or aliasing in our period determination. Fits to the positions and constant angular velocity of the motion both indicate a near-circular orbit. Figure 3 further enhances the satellite's visibility; it is a composite of deconvolved single frames. The provisional IAU designation for the satellite is S/1998(45)1 (ref. 15).

We can reject non-satellite explanations for our observations. No known artefact in the point-spread-function (PSF) produces a point source at different position angles and separations in deep 4-minute exposures. This object was stable in position and brightness, unlike other semi-persistent similar-appearing bright spots in the diffraction pattern. None of the PSF stars (used for calibration of the instrumental profile) had similar secondary point-sources located near the $0.5\text{--}0.8''$ separations observed here. The satellite, although 6 magnitudes fainter than the primary, was at times diffraction-limited, indicating that the source of the signal was external to the telescope and optical system. The companion has been imaged in three near-infrared filters (J at $1.2 \mu\text{m}$ wavelength, H at $1.6 \mu\text{m}$, and K' at $2.1 \mu\text{m}$). Background stars are easily eliminated because of their rapid motion with respect to the tracked asteroid. No matches were found with the positions of known Solar System objects.

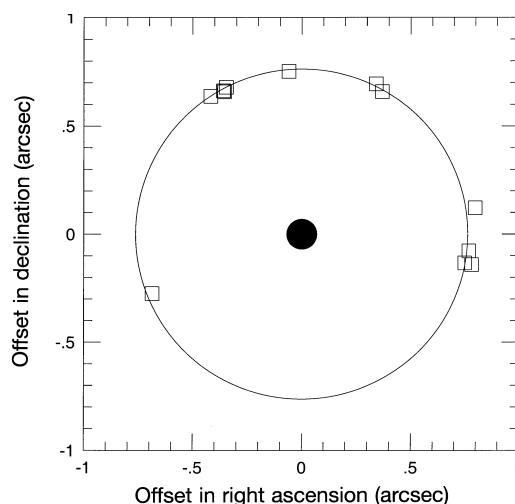


Figure 2 How the orbit of the satellite around 45 Eugenia would appear if it were observed face-on. Observations have been rotated by 34° clockwise and deprojected with respect to line-of-sight by 46° . Our orbit-pole solution is ecliptic (longitude, latitude) = $(135^\circ, -19^\circ)$, which puts the orbit close to Eugenia's equatorial plane. Clusters of points are separate detections within a single night. The central filled circle is scaled to the mean size of Eugenia.

Table 1 Properties of 45 Eugenia and targets of spacecraft fly-bys

Asteroid	Type	Albedo	Diameter (km)	Density (g cm ⁻³)	Satellite?	Family member?
45 Eugenia	F (C)	0.04	215	1.2	Yes	Yes ^{30,31}
951 Gaspra	S	0.23	12	ND	No	Uncertain ³²
243 Ida	S	0.21	31	2.6	Yes	Yes
253 Mathilde	C	0.04	53	1.3	No	No ³²
433 Eros	S	0.27	18	2.5	No	No

ND, not determined.

Follow-up observations on 4 January 1999 UT also show the object, in a position consistent with the orbit we derive here.

As one of the 25 largest asteroids, Eugenia is well observed. It is a 'fast-rotator' with a spin period of 5.7 hours and, like other C-like asteroids, is exceedingly black (albedo 0.040; ref. 4). It is classified as an F-type (similar to C-type), with a large-amplitude light curve, implying an elongated shape, perhaps like a jacobian ellipsoid (with aspect ratios $a/b = 1.33$, $b/c = 1.65$, where a , b , c are the axes of the triaxial ellipsoid)¹⁶, suggesting that it may have rubble-pile structure.

The satellite's orbit is inclined 46° to the line of sight. Taylor *et al.*¹⁷ showed that Eugenia has a retrograde rotation, but with a two-position ambiguity in the pole position (that is, the direction of the spin angular momentum vector). Drummond *et al.*¹⁶ rejected one of those two pole positions, favouring only that at ecliptic (longitude, latitude) = (127°, -44°). Taylor *et al.* and subsequent authors agree with this pole to within 20°. If we assume that the satellite's orbit is in Eugenia's equatorial plane, then our data are entirely consistent with the pole proposed by Drummond *et al.*, but not with the second pole (286°, -26°) of Taylor *et al.* We suggest that the satellite is roughly in the equatorial plane, and that it is in a prograde orbit (same sense as primary spin), having an orbital pole within 10° of (135°, -19°). A prograde orbit is preferred for a satellite formed from impact-generated orbital debris^{7,18}. A retrograde orbit, however, is more stable against perturbing effects of the non-uniform gravitational field of an oblate primary^{19,20}. An orbit with an opposite sense to the asteroid's orbital motion around the Sun (as is the case here) is more stable against the effects of solar tides²¹. The derived orbit for S/1998(45)1 is stable against both of these perturbations, because it is close enough to Eugenia to minimize solar tidal effects, but far enough to be unaffected by Eugenia's shape.

We find the orbital period to be 4.691 ± 0.004 days. The long-axis separation of S/1998(45)1 from Eugenia was 0.76" (21.8 pixels) on

1 November, when it was 2.143 astronomical units (AU) from Earth, yielding a semimajor axis of $1,190 \pm 30$ km. The separation and period give a gravitational parameter (gravitational constant times the mass of Eugenia) GM of $0.40 \text{ km}^3 \text{ s}^{-2}$. We adopt the mean IRAS (ref. 4) diameter of 215 ± 4 km for Eugenia, yielding a bulk density of Eugenia of 1.2 g cm^{-3} . The formal IRAS errors would imply an error in our density estimate of $<0.2 \text{ g cm}^{-3}$. But the IRAS error does not account for a very non-spherical Eugenia; perhaps the IRAS observations were taken when the asteroid's cross-section was larger than the mean²². If so, the density could be as high as 1.8 g cm^{-3} . If instead of the IRAS errors, we use the maximum deviation expected, based on the asteroid's aspect ratio, conservative error bars are $\pm 0.6 \text{ g cm}^{-3}$, and are dominated by the uncertainty in Eugenia's size, not in the satellite's orbital parameters. Standard ground-based techniques can readily improve knowledge of Eugenia's size. From the brightness ratio of Eugenia to its satellite (6.14 ± 0.1 mag in the H-band, including uncertainties in Eugenia's projected diameter, and assuming the same albedo for both), we estimate the diameter of S/1998(45)1 as 13 ± 3 km.

Although not yet thoroughly examined, data from the other 25 asteroids that we observed do not show such prominent evidence for satellites as Eugenia. Other searches, both ground-based^{9,23} and from the Hubble Space Telescope²⁴, have found no satellites in the more limited parameter space they sampled. Data from close encounters of four asteroids by spacecraft—951 Gaspra²⁵, 243 Ida¹⁻³, 253 Mathilde^{5,13} and 433 Eros^{14,26}—reveal only one satellite. Radar imaging of asteroids show apparent contact binaries, but no small satellites. Light curves of near-Earth asteroids suggest that some may be similar-sized binary pairs^{27,28}. Although our discovery of Eugenia's satellite demonstrates that Dactyl was not a fluke, there are growing indications that asteroidal satellites are not common.

Hypotheses for Dactyl's origin^{12,29} generally invoke retention of fragments in a catastrophically disruptive collision, of the type believed to form asteroid families. These studies, coupled with our detection of S/1998(45)1, suggest an emerging correlation between family membership and formation of a satellite (see Table 1). Alternatively, satellites may be formed from orbiting ejecta after a non-catastrophic oblique impact⁷. Eugenia's non-spherical shape enhances the chances that ejecta will avoid re-impact, even on keplerian trajectories, and hence of establishing orbital debris that might accrete into a satellite. Weidenschilling *et al.*⁷ state that such satellites would be small relative to the primary, and would be in prograde orbits close (within a few radii initially, but evolving outward) to a rapidly-spinning primary. Eugenia and

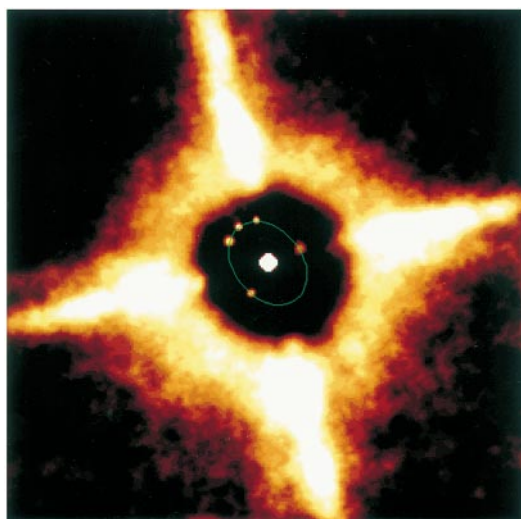


Figure 3 Orbit of the satellite around 45 Eugenia as observed. This image is a composite of individual observations taken on five separate nights (clockwise from top: 6, 7, 9, 10 and 1 November 1998 UT). North is up and east is to the left. The images are deconvolved, and the intensity of Eugenia is suppressed to enhance sharpness and clarity. The green line depicts the orbit as seen on the plane of the sky. The large, bright cross pattern results from scattering by the telescope's secondary mirror support. The variable size and brightness of the satellite reflects the variable size of the AO FWHM and is not physical.

its satellite meet these conditions. It is even possible that Eugenia's rapid spin and resulting unusual shape may be due to the collision that formed S/1998(45)1.

Our most intriguing result is Eugenia's low density. In Table 1 we compare Eugenia with other C- and S-type asteroids with directly determined densities. There may be two groupings of densities, one near 1.2 g cm^{-3} and one near 2.5 g cm^{-3} , with different associated albedos. Mathilde's low density may result from a loosely packed, porous, rubble pile of materials of rock-like intrinsic densities of $\sim 3.5 \text{ g cm}^{-3}$ (ref. 5). Mathilde and Eugenia are clearly structurally or compositionally distinct from the S-types Ida and Eros; but Ida may also have a rubble-pile structure²⁹, despite its higher density. An object as large as Eugenia would have higher internal pressures than Mathilde; it is doubtful whether such a body—made of inherently weak carbonaceous materials characteristic of its surface spectrum and albedo—could maintain the high porosity required to explain its low bulk density. Another possibility is that Mathilde, Eugenia and other C-like asteroids may be composed primarily of water ice, perhaps as remnants of burned-out comet-like bodies that have crusts of less-volatile carbonaceous material which result in their low albedos. □

Received 1 July; accepted 8 September 1999.

- Belton, M. J. S. *et al.* Bulk density of asteroid 243 Ida from the orbit of its satellite Dactyl. *Nature* **374**, 785–788 (1995).
- Belton, M. J. S. *et al.* The discovery and orbit of 1993(243)1 Dactyl. *Icarus* **120**, 185–199 (1996).
- Chapman, C. R. *et al.* Discovery and physical properties of Dactyl, a satellite of asteroid 243 Ida. *Nature* **374**, 783–785 (1995).
- Telesco, E. F., Veeder, G. J., Fowler, J. W. & Chillemi, J. R. *The IRAS Minor Planet Survey—Final Report* (PL-TR-92-2049, Phillips Laboratory, Hanscom Air Force Base, MA, 1992).
- Veverka, J. *et al.* NEAR's flyby of 253 Mathilde: images of a C-asteroid. *Science* **278**, 2109–2114 (1997).
- Yeomans, D. K. *et al.* Estimating the mass of asteroid 253 Mathilde from tracking data during the NEAR flyby. *Science* **278**, 2106–2109 (1997).
- Weidenschilling, S. J., Paolicchi, P. & Zappala, V. in *Asteroids II* (eds Binzel, R., Gehrels, T. & Matthews, M.) 643–658 (Univ. Arizona Press, Tucson, 1989).
- Van Flandern, T. C., Telesco, E. F. & Binzel, R. P. in *Asteroids* (ed. Gehrels, T.) 443–465 (Univ. Arizona Press, Tucson, 1979).
- Gehrels, T., Drummond, J. D. & Levenson, N. A. The absence of satellites of asteroids. *Icarus* **70**, 257–263 (1990).
- Chauvineau, B. & Mignard, F. Dynamics binary asteroids II. Jovian perturbations. *Icarus* **87**, 377–390 (1990).
- Botke, W. F. Jr & Melosh, H. J. Formation of asteroidal satellites and doublet craters by planetary tidal forces. *Nature* **381**, 51–53 (1996).
- Durda, D. D. The formation of asteroidal satellites in catastrophic collisions. *Icarus* **120**, 212–219 (1996).
- Merline, W. J. *et al.* NEAR's encounter with 253 Mathilde: Search for satellites. *Proc. Lunar Planet. Sci. Conf. XXIX*, abstract no. 1954 (1998).
- Merline, W. J. *et al.* Search for satellites around asteroid 433 Eros from NEAR flyby imaging. *Proc. Lunar Planet. Sci. Conf. XXX*, abstract no. 2055 (1999).
- Merline, W. J. *et al.* in S/1998(45)1. *IAU Circ. No.* 7129, (1999).
- Drummond, J. D., Weidenschilling, S. J., Chapman, C. R. & Davis, D. R. Photometric geodesy of main-belt asteroids. II. Analysis of lightcurves for poles and shapes. *Icarus* **73**, 314–323 (1988).
- Taylor, R. C., Birch, P. V., Pospieszalska-Surdej, A. & Surdej, J. Asteroid 45 Eugenia: lightcurves and pole orientation. *Icarus* **73**, 314–323 (1988).
- Durda, D. D. & Geissler, P. E. The formation of asteroidal satellites in large cratering collisions. *Bull. Am. Astron. Soc.* **28**, 1101 (1996).
- Chauvineau, B., Farinella, P. & Mignard, F. Planar orbits about a triaxial body: Applications to asteroidal satellites. *Icarus* **105**, 370–384 (1993).
- Sheeres, D. J. Dynamics about uniformly rotating tri-axial ellipsoids. *Icarus* **110**, 225–238 (1994).
- Hamilton, D. P. & Burns, J. A. Orbital stability zones about asteroids. *Icarus* **92**, 118–131 (1991).
- Harris, A. W., Young, J. W., Bowell, E. & Tholen, D. J. Asteroid lightcurve observations from 1981–1983. *Icarus* (in the press).
- Grady, J. & Flynn, L. A search for satellites and dust belts around asteroids: negative results. *Proc. Lunar Planet. Sci. Conf. XIX*, 405–406 (1988).
- Storrs, A. *et al.* Imaging observations of asteroids with Hubble Space Telescope. *Icarus* **137**, 260–268 (1999).
- Belton, M. J. S. *et al.* Galileo encounter with 951 Gaspra: First pictures of an asteroid. *Science* **257**, 1647–1652 (1992).
- Veverka, J. *et al.* Imaging of asteroid 433 Eros during NEAR's flyby reconnaissance. *Science* **285**, 562–564 (1999).
- Mottola, S. & Lahulla, F. Mutual eclipse events in binary asteroid system 1996 FG3: observations and a numerical model. *Asteroids, Comets, Meteors '99*, paper 24.03 (1999); (http://scorpio.tn.cornell.edu/ACM/web_abs.html).
- Pravec, P., Wolf, M., Sarounova, L. Occultation/eclipse events in binary asteroid 1991 VH. *Icarus* **133**, 79–88 (1998).
- Davis, D. R., Chapman, C. R., Durda, D. D., Farinella, P. & Marzari, F. The formation and collisional/dynamical evolution of the Ida/Dactyl system as part of the Koronis family. *Icarus* **120**, 220–230 (1996).
- Chapman, C. R., Paolicchi, P., Zappala, V., Binzel, R. P. & Bell, J. F. in *Asteroids II* (eds Binzel, R., Gehrels, T. & Matthews, M.) 386–415 (Univ. Arizona Press, Tucson, 1989).

- Bus, S. J. *Compositional Structure in the Asteroid Belt: Results of a Spectroscopic Survey*. Thesis, Massachusetts Inst. Technol. (1999).
- Zappala, V., Bendjoya, Ph., Cellino, A., Farinella, P. & Froeschle, C. Asteroid families: Search of a 12,487-asteroid sample using two different clustering techniques. *Icarus* **116**, 291–314 (1995).
- Roddi, F. Curvature sensing and compensation: a new concept in adaptive optics. *Appl. Opt.* **27**, 1223–1225 (1988).
- Roddi, F., Northcott, M. & Graves, J. E. A simple low-order adaptive optics system for near infrared applications. *Publ. Astron. Soc. Pacif.* **103**, 131–149 (1991).
- Rigaut, F. *et al.* Performance of the Canada-France-Hawaii Telescope adaptive optics bonnet. *Publ. Astron. Soc. Pacif.* **110**, 152–164 (1998).

Acknowledgements

W.J.M., L.M.C., C.D., F.R., F.M. and G.D. are visiting astronomers at the Canada-France-Hawaii Telescope (CFHT), operated by the National Research Council of Canada, the Centre National de la Recherche Scientifique of France and the University of Hawaii. This work was done using observing time allocated through the University of Hawaii and granted by the CFHT Corporation, and was supported by funding from NASA and the NSF. We thank the staff of CFHT, particularly J.-L. Beuzit, for their logistical support; M. Northcott and J. E. Graves for development of a prototype instrument; and W. Colwell for assistance with data analysis tools.

Correspondence and requests for materials should be addressed to W.J.M. (e-mail: merline@boulder.swri.edu).

Preparing topological states of a Bose–Einstein condensate

J. E. Williams & M. J. Holland

JILA and Physics Department, University of Colorado, Boulder, Colorado 80309-0440, USA

Observations of Bose–Einstein condensates—macroscopic populations of ultracold atoms occupying a single quantum state—in dilute alkali-metal and hydrogen gases have stimulated a great deal of research into the statistical physics of weakly interacting quantum degenerate systems^{1,2}. Recent experiments offer a means of exploring fundamental low-temperature physics in a controllable manner. A current experimental goal in the study of trapped Bose gases is the observation of superfluid-like behaviour, analogous to the persistent currents seen in superfluid liquid helium which flow without observable viscosity. The ‘super’ properties of Bose-condensed systems occur because the macroscopic occupation of a quantized mode provides a stabilizing mechanism that inhibits decay through thermal relaxation³. Here we show how to selectively generate superfluid vortex modes with different angular momenta in a Bose–Einstein condensate. Our approach involves solving the time-dependent equation of motion of a two-component condensate with strongly coupled internal atomic states, as recently investigated experimentally^{4,5}. The generation of vortices relies on the coupling between the states (achieved by applying an electromagnetic field), combined with mechanical rotation of the trapping potentials which confine the condensate.

Since 1995, when Bose–Einstein condensation in a dilute atomic gas was first observed^{6–8}, experimenters have sought a method to create a vortex in this system. In a typical experiment, around one million atoms are trapped in a magnetic harmonic potential and cooled below the critical temperature so that condensation occurs into the lowest energy quantized mode. In the usual case, this ground state has no circulation. One proposed scheme for preparing the condensate in a vortex mode^{9–17} is to distort the confining potential and mechanically rotate the trap during the cooling process. In this way, the lowest energy mode may be engineered to be circulating about the axis of symmetry. Such an approach is in

direct analogy with experiments on vortices in superfluid helium—the asymmetry of the harmonic trap for the atomic gas plays the role of surface roughness of a rotating vessel. Although conceptually this method appears promising for vortex generation in a trapped gas, so far technical difficulties have precluded its successful implementation.

Instead of having the system condense into a vortex mode, an alternative approach is to allow the atoms to condense into the usual ground state and then dynamically generate the vortex from the non-rotating condensate. Several theoretical proposals have been made along these lines which utilize the interaction between the atoms and a specific laser field consisting of a beam of photons with non-zero orbital angular momentum^{18–20}.

The method we present here makes use of both of the techniques mentioned—mechanical rotation and the coupling of internal states using an electromagnetic field. It is motivated by the recent observations demonstrating the cyclic twisting and untwisting of the order parameter of a gaseous condensate of rubidium atoms⁴. These recent experiments have suggested the possibility for harnessing the interplay between the internal and motional dynamics in order to fundamentally alter the topological structure of the order parameter of the condensate⁵.

We treat the internal structure of the condensed atoms as a two-level system, as illustrated in Fig. 1. The two states are confined in separate axially symmetric harmonic oscillator potentials with the same trap frequency ω_0 . As shown in Fig. 1a, the trap centres are spatially offset by a distance r_0 and are rotated about the symmetry axis at a frequency of ω_r . Simultaneously an electromagnetic field is applied that couples the two internal atomic hyperfine states causing the atoms to coherently cycle between levels, as illustrated in Fig. 1b. There are two parameters that characterize the coupling: the detuning and the power. The detuning δ denotes the mismatch of the frequency of the coupling electromagnetic field to the frequency different between the two internal atomic states. The power is characterized by the Rabi frequency Ω , which is the rate at which population would oscillate between the two states if δ were zero. When δ is larger than Ω , so that the drive is off resonance, the population oscillations have small amplitude and occur at the effective Rabi frequency $\Omega_{\text{eff}} = \sqrt{\Omega^2 + \delta^2}$.

The dynamical evolution of the condensate state $|\psi(t)\rangle$ for a gas of atoms in two different hyperfine states is governed by a nonlinear Schrödinger equation, known as the Gross–Pitaevskii equation, generalized to treat the internal state coupling⁵

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = \left[\hat{H}_0 \otimes \hat{1} + \hat{1} \otimes \frac{\hbar}{2} (\Omega \hat{\sigma}_x + \delta \hat{\sigma}_z) + \hat{H}_1 \otimes \hat{\sigma}_z \right] |\psi(t)\rangle \quad (1)$$

where $\{\hat{1}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}$ is the set of standard Pauli spin operators and for clarity we have used the tensor product $\otimes$ to explicitly separate the

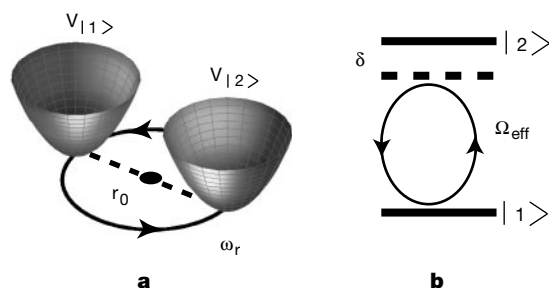


Figure 1 Method for creating a vortex. **a**, The two traps V_{11} and V_{12} are rotated in the x – y plane about the z axis at a frequency ω_r . **b**, Transitions are simultaneously driven between the two internal states at the effective Rabi frequency Ω_{eff} . A vortex mode possessing one unit of angular momentum can be prepared if $\omega_r \approx \Omega_{\text{eff}}$.

spatial and internal operators. The free hamiltonian $\hat{H}_0$ describes the motional dynamics of the atoms in a stationary harmonic trap centred on the symmetry axis

$$\hat{H}_0(\mathbf{r}) = -\frac{\hbar^2}{2m} \nabla^2 + \frac{1}{2} m \omega_0^2 r^2 + U_0 n(\mathbf{r}) \quad (2)$$

where m is the atomic mass. The term that makes the system nonlinear is the mean-field interaction energy, which depends on the local density of condensate atoms $n(\mathbf{r})$. The coefficient $U_0 = 4\pi\hbar^2 a/m$ is proportional to the scattering length of a binary collision a . The effect of displacing the trap centres and rotating them about the symmetry axis is described by

$$\hat{H}_1(\mathbf{r}, t) = \kappa [f(\mathbf{r}) \cos(\omega_r t) + g(\mathbf{r}) \sin(\omega_r t)] \quad (3)$$

where κ is a coupling coefficient and $f(\mathbf{r})$ and $g(\mathbf{r})$ are functional prefactors. The explicit form of $\hat{H}_1$ determines the symmetry of the quantum state being prepared. To create a vortex state with one unit of angular momentum $\kappa = m\omega_0^2 r_0$, $f(\mathbf{r}) = x$ and $g(\mathbf{r}) = y$ in cartesian coordinates, corresponding to the scheme depicted in Fig. 1a. More general forms for f and g will be considered later, with equation (3) then providing the definition of a generalized κ and ω_r .

Before solving this problem in detail, we first present an intuitive picture of the underlying physics. The vortex state we wish to create has unit circulation corresponding to a macroscopically occupied single particle wavefunction with quantized angular momentum $\langle \hat{L}_z \rangle / \hbar = 1$. In order to see why our scheme couples a non-rotating condensate to this state, consider the frame co-rotating with the trap centres at angular frequency ω_r so that $\hat{H}_1$ becomes time-independent. The free hamiltonian in the co-rotating frame is given by²¹ $\hat{H}_0 - \omega_r \hat{L}_z$. The energy of the vortex with one unit of angular momentum is therefore shifted by $\hbar\omega_r$ in the rotating frame relative

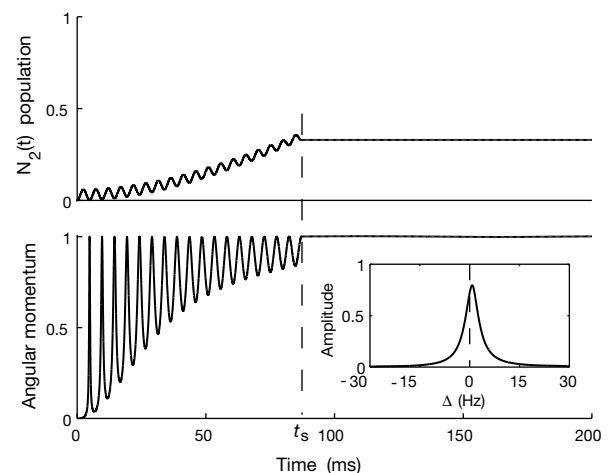


Figure 2 Dynamical evolution to a vortex. These are results of a numerical integration of equation (1), with the condensate initially in the non-rotating ground state and in the internal state $|1\rangle$. The coupling drive is turned on at time $t = 0$, and is turned off at time $t = t_s$ by setting both Ω and r_0 to zero. The top graph shows the fractional population of atoms in the $|2\rangle$ internal state. The small-amplitude rapid oscillations correspond to the cycling between internal levels due to the off-resonant coupling. The gradual rise of this line is due to coupling from the ground state to the vortex mode caused by the drive H_1 in equation (3). The bottom graph shows the angular momentum of the $|2\rangle$ state, in units of Planck's constant $\hbar$. The rise and fall of this curve corresponds to a rapid cycling of the $|2\rangle$ atoms between the non-rotating condensate and the vortex. Once during each Rabi cycle, the angular momentum approaches unity and at that time the $|2\rangle$ state wavefunction approaches a pure vortex mode. The inset shows the maximum amplitude of population transfer to the vortex as a function of the trap rotation frequency ω_r , with $\Delta = \Omega_{\text{eff}} - \omega_r$. The various parameters used in this calculation are: $\omega_0 2\pi = 10$ Hz, $\delta 2\pi = 200$ Hz, $\omega_r 2\pi = 205.4$ Hz, $N = 8 \times 10^5$ atoms, m is the mass of the ^{87}Rb atom, $a = 5.5$ nm (ref. 25), and $\Omega 2\pi = 50$ Hz and $r_0 = 1.7 \mu\text{m}$ for $t < t_s$.

to its value in the laboratory frame. When this energy shift compensates for both the energy mismatch $\hbar\delta$ of the internal coupling field and the small chemical potential difference between the vortex and non-rotating condensate, resonant transfer of population may take place.

For this picture to be valid, the following inequalities must be satisfied

$$\omega_r \gg \omega_0, \quad \delta \gg \Omega, \quad \omega_r \approx \Omega_{\text{eff}} \quad (4)$$

which also imply $\Omega_{\text{eff}} \approx \delta$. The first inequality splits significantly the energies of states of different angular momenta in the rotating frame. This point is crucial for experimental implementation, as it allows the vortex to be generated rapidly and is a feature absent from previous dynamic coupling schemes^{18–20}. The weak coupling limit given by the second inequality allows the resonance condition $\omega_r \approx \Omega_{\text{eff}}$ to select energetically the desired state with high fidelity. Even though we require a large value for ω_r , one may not use a simple equivalent time-averaged potential for the effect of rotating the trap centres without including the internal atomic state dynamics in the time-averaging, which occur on the same timescale.

In order to visualize the system dynamics, we present results from a numerical integration of the Gross–Pitaevskii equation²². Here we treat the system in two dimensions in the x – y plane perpendicular to the symmetry axis. In Fig. 2 we plot as a function of time the population fraction and the angular momentum per atom of the $|2\rangle$ state. We initialize the system with all of the atoms in the $|1\rangle$ internal level and in the mean-field ground state of the trap. There are two striking features in Fig. 2. First, even though we are driving the internal states far from the atomic resonance, a significant fraction of the population gets transferred to the $|2\rangle$ state over a time long compared to the fast Rabi oscillations at Ω_{eff} . Second, we see that the angular momentum of state $|2\rangle$ oscillates rapidly at the frequency Ω_{eff} .

Our key idea is that by turning off the coupling at a precise time, $t = t_s$, on a given Rabi cycle, the $|2\rangle$ state can be prepared to have unit angular momentum. The maximum possible population

transfer to the vortex state using this scheme obeys a lorentzian response curve as ω_r is varied near Ω_{eff} exhibiting a narrow resonance. This is shown in the inset of Fig. 2, where $\Delta = \Omega_{\text{eff}} - \omega_r$, and will be discussed in more detail later. In Fig. 3 we show a ‘snapshot’ of the densities and phases of the two components at time $t = 200$ ms. The snapshot illustrates the preparation of a high-quality vortex in state $|2\rangle$, with the $|1\rangle$ state providing a natural ‘pinning’ mechanism that stabilizes the vortex core by providing a repulsive mean-field barrier along the symmetry axis. Consequently the size of the vortex core is determined by the spatial density profile of the non-rotating state, and not by the natural healing length.

An intriguing property of our state preparation scheme is that the direction of circulation of the vortex can be opposite to that of the rotating trap centres; changing the sign of the detuning δ with the direction of rotation of the trap centres fixed causes the vortex to rotate in the opposite direction. This is most easily seen by again considering the frame co-rotating with the potentials at frequency ω_r . Vortices with opposite circulations experience opposite energy shifts in transforming to the rotating frame, and therefore require opposite signs of detuning in order to achieve resonant coupling.

Although the numerical calculations we present here are full solutions of equation (1), an accurate and greatly simplified model may be derived using the separation of timescales given by the inequalities in equation (4). This allows us to gain insight into the physical mechanisms giving rise to the behaviour of this system. If we time-average over rapidly oscillating terms in order to study the long-timescale behaviour, we arrive at an approximate solution for the system dynamics, which has the simple form²³

$$|\psi(t)\rangle = [a(t)c_0(t)|\phi_0\rangle + b(t)c_n(t)|\phi_n\rangle]|1\rangle + [b(t)c_0(t)|\phi_0\rangle + a^*(t)c_n(t)|\phi_n\rangle]|2\rangle \quad (5)$$

Here $|\phi_0\rangle$ and $|\phi_n\rangle$ are the spatial parts of the state, with $|\phi_0\rangle$ being the ground state and $|\phi_n\rangle$ being the target state—the unit vortex for the

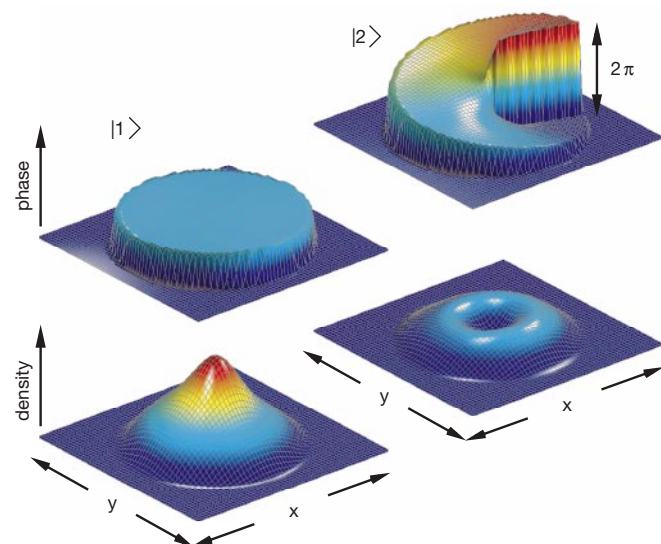


Figure 3 Unit vortex preparation. Shown are the calculated densities and phases of the two states $|1\rangle$ and $|2\rangle$, at time $t = 200$ ms indicated in Fig. 2. At this time, about one-third of the atoms are in the $|2\rangle$ state, which is in a pure vortex mode with unit angular momentum. A characteristic feature of a vortex mode is the $2n\pi$ phase wrap about the core, where n is an integer that is equal to unity in this case. An attractive feature of this preparation scheme is that the $|1\rangle$ atoms residing in the core region provide a natural pinning mechanism for the vortex due to the mean-field repulsion.

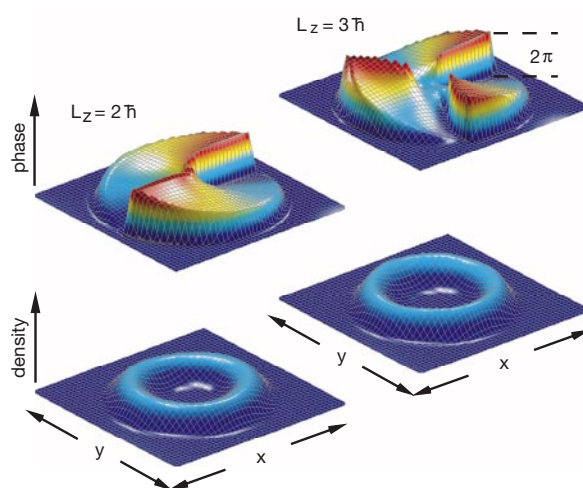


Figure 4 Double and triple vortex preparation. Shown are the densities and phases modulo 2π of the $|2\rangle$ state after a dynamical evolution similar to that shown in Fig. 2, but with different symmetries of the drive H_1 . In the case of $\langle \hat{L}_z \rangle / \hbar = 2$, $f = x^2 - y^2$ and $g = 2xy$, while for $\langle \hat{L}_z \rangle / \hbar = 3$, $f = x^3 - 3xy^2$ and $g = 3x^2y - y^3$. In both cases, the system was evolved from the same initial condition as that for the calculation described in Fig. 2, with about one-third of the atoms in the $|2\rangle$ state at the time t_s . The values taken for the various parameters were the same, except for the trap rotation frequency, which was $\omega_r 2\pi = 204.3$ Hz for the $\langle \hat{L}_z \rangle / \hbar = 2$ case, and $\omega_r 2\pi = 200.2$ Hz for the $\langle \hat{L}_z \rangle / \hbar = 3$ case.

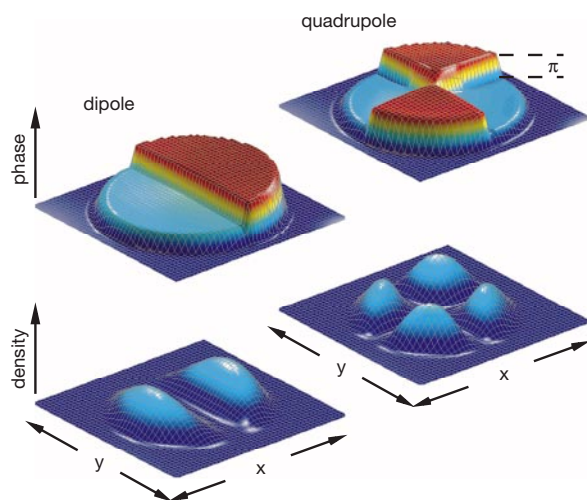


Figure 5 Dipole and quadrupole preparation. Different symmetries were constructed using specific forms for the drive H_1 in order to prepare the $|2\rangle$ state in non-circulating modes. Instead of having a $2n\pi$ phase wrap corresponding to a current flowing around a central core, these modes have regions of constant phase separated by a discontinuous jump of π where the wavefunction changes sign. To generate the dipole mode, we used $f = x$ and $g = 0$ with $\omega_c 2\pi = 205.4$ Hz; while for the quadrupole, we used $f = xy$ and $g = 0$ with $\omega_c 2\pi = 204.3$ Hz.

case considered in Fig. 3, determined by the self-consistent lowest energy solutions of the two-component Gross–Pitaevskii equation⁵. For given populations of the two atomic states, the energy is minimized using a steepest-descent algorithm with the constraint that $|\phi_0\rangle$ is the nodeless ground state and $|\phi_n\rangle$ possesses the symmetry imposed by $\hat{H}_1$, which for the unit vortex is a 2π topological phase wrap on any closed loop containing the core. The rapidly varying coefficients in equation (5) are found to be given by $a(t) = \cos(\Omega_{\text{eff}} t/2) - i(\delta/\Omega_{\text{eff}}) \sin(\Omega_{\text{eff}} t/2)$ and $b(t) = -i(\Omega/\Omega_{\text{eff}}) \sin(\Omega_{\text{eff}} t/2)$. The slow coupling of the spatial states is given by c_0 and c_n , which satisfy

$$i\hbar \begin{pmatrix} \dot{c}_0 \\ \dot{c}_n \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -(\epsilon_n - \epsilon_0 - \hbar\Delta) & \beta\langle\phi_n|f + ig|\phi_0\rangle \\ \beta\langle\phi_0|f - ig|\phi_n\rangle & (\epsilon_n - \epsilon_0 - \hbar\Delta) \end{pmatrix} \begin{pmatrix} c_0 \\ c_n \end{pmatrix} \quad (6)$$

Here $\beta = \kappa\Omega/|\delta|$ is the coupling coefficient, f and g are defined in equation (3), and ϵ_0 and ϵ_n are the energy eigenvalues of $|\phi_0\rangle$ and $|\phi_n\rangle$, respectively. The timescale for state preparation is determined by β and the overlap matrix element $\langle\phi_n|f + ig|\phi_0\rangle$. These two parameters also determine the width of the response lorentzian that is shown in the inset of Fig. 2, with the maximum occurring not at zero frequency but displaced by an amount according to the splitting between the eigenenergies $\epsilon_n - \epsilon_0$; typically a small fraction of ω_0 .

The symmetry of the coupling is determined by $f + ig$, which has the form $x + iy$ as previously stated for the case of the unit vortex. In order to produce a vortex with n units of angular momentum (that is $\langle\hat{L}_z\rangle/\hbar = n$), $f + ig$ should take on the form $(x + iy)^n$. So for $n = 2$, we must construct an $\hat{H}_1$ in which $f = x^2 - y^2$ and $g = 2xy$, corresponding to a saddle potential rotating at $\omega_c/2$. In Fig. 4, we illustrate vortex generation with two and three units of angular momentum.

The generalization of our scheme for the preparation of macroscopic quantum states of arbitrary symmetry is straightforward. For example, to generate a mode with a dipole symmetry one superimposes a left- and right-circulating vortex so that $f = \text{Re}\{x + iy\}$ and $g = 0$. A quadrupole symmetry would be generated by $f = \text{Im}\{(x + iy)^2\}$ and $g = 0$. The dynamical state preparation of these two examples is illustrated in Fig. 5.

We do not anticipate collisions to play a significant role on the short timescale for vortex generation shown in Fig. 2, and we have implicitly neglected them in our zero temperature theory. However, the system is thermodynamically unstable on long timescales if the rotating drive $\hat{H}_1$ is left on, as atoms can access states of higher and higher angular momentum through collisional relaxation. Even after the vortex state has been generated and the drive is turned off, collisions will be important in determining the stability of

persistent currents where the topological structure of the two-component order parameter is known to be crucial^{3,12}.

Note: Since this Letter was submitted, vortices in a two-component Bose–Einstein condensate have been created and observed using this scheme²⁴. □

Submitted 21 June; accepted 10 September 1999.

1. Cornell, E. A., Ensher, J. R. & Wieman, C. E. Experiments in dilute atomic Bose–Einstein condensation. In *Proc. 1998 Enrico Fermi Summer School on Bose–Einstein Condensation* (eds Inguscio, M., Stringari, S. & Wieman, C. E.) (SIF, Bologna, in the press); also as preprint cond-mat/9903109 at <http://xxx.lanl.gov> (1999).
2. Ketterle, W., Durfee, D. S. & Stamper-Kurn, D. M. Making, probing and understanding Bose–Einstein condensates. In *Proc. 1998 Enrico Fermi Summer School on Bose–Einstein condensation* (eds Inguscio, M., Stringari, S. & Wieman, C. E.) (SIF, Bologna, in the press); also as preprint cond-mat/9904034 at <http://xxx.lanl.gov> (1999).
3. Leggett, A. in *The New Physics* (ed. Davies, P.) 268–288 (Cambridge Univ. Press, New York, 1989).
4. Matthews, M. R. *et al.* Watching a superfluid untwist itself: Recurrence of Rabi oscillations in a Bose–Einstein condensate. *Phys. Rev. Lett.* (in the press); also as preprint cond-mat/9906288 at <http://xxx.lanl.gov> (1999).
5. Williams, J., Walser, R., Cooper, J., Cornell, E. A. & Holland, M. Excitation of an antisymmetric collective mode in a strongly coupled two-component Bose–Einstein condensate. Preprint cond-mat/9904399 at <http://xxx.lanl.gov> (1999).
6. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
7. Davis, K. B. *et al.* Bose–Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
8. Bradley, C. C., Sackett, C. A., Tollet, J. J. & Hulet, R. G. Evidence of Bose–Einstein condensation in an atomic gas with attractive interactions. *Phys. Rev. Lett.* **75**, 1687–1690 (1995); Erratum. *Phys. Rev. Lett.* **79**, 1170–1170 (1997).
9. Butts, D. A. & Rokhsar, D. S. Predicting signatures of rotating Bose–Einstein condensates. *Nature* **397**, 327–329 (1999).
10. Rokhsar, D. S. Vortex stability and persistent currents in trapped Bose gases. *Phys. Rev. Lett.* **79**, 2164–2167 (1997).
11. Dalfovo, F. & Stringari, S. Bosons in anisotropic traps: Ground state and vortices. *Phys. Rev. A* **53**, 2477–2485 (1996).
12. Ho, T.-L. & Shenoy, V. B. Local spin-gauge symmetry of the Bose–Einstein condensates in atomic gases. *Phys. Rev. Lett.* **77**, 2595–2599 (1996).
13. Barenghi, C. F. Vortex waves in a cloud of Bose–Einstein-condensed, trapped alkali-metal atoms. *Phys. Rev. A* **54**, 5445–5446 (1996).
14. Fetter, A. L. Vortex stability in a trapped Bose condensate. *J. Low Temp. Phys.* **113**, 189–194 (1998).
15. Feder, D. L., Clark, C. W. & Schneider, B. I. Vortex stability of interacting Bose–Einstein condensates confined in anisotropic harmonic traps. *Phys. Rev. Lett.* **82**, 4956–4959 (1999).
16. Caradoc-Davis, B. M., Ballagh, R. J. & Burnett, K. Coherent dynamics of vortex formation in trapped Bose–Einstein condensates. *Phys. Rev. Lett.* **83**, 895–898 (1999).
17. Dodd, R. J., Burnett, K., Edwards, M. & Clark, C. W. Excitation spectroscopy of vortex states in dilute Bose–Einstein condensed gases. *Phys. Rev. A* **56**, 587–590 (1997).
18. Marzlin, K.-P. & Zhang, W. Vortex coupler for atomic Bose–Einstein condensates. *Phys. Rev. Lett.* **79**, 4728–4731 (1997).
19. Bolda, E. L. & Walls, D. F. Creation of vortices in a Bose–Einstein condensate by a Raman technique. *Phys. Lett. A* **246**, 32–36 (1998).
20. Dum, R., Cirac, J. I., Lewenstein, M. & Zoller, P. Creation of dark solitons and vortices in Bose–Einstein condensates. *Phys. Rev. Lett.* **80**, 2972–2975 (1998).
21. Lifshitz, E. M. & Pitaevskii, L. P. *Statistical Physics* (Pergamon, Oxford, 1980).
22. Holland, M. J., Jin, D. S., Chiofalo, M. L. & Cooper, J. Emergence of interaction effects in Bose–Einstein condensation. *Phys. Rev. Lett.* **78**, 3801–3805 (1997).
23. Williams, J. E. *The Preparation of Topological Modes in a Strongly-Coupled Two-Component Bose–Einstein Condensate*. Thesis, Univ. Colorado (1999).

24. Matthews, M. R. *et al.* Vortices in a Bose-Einstein condensate. *Phys. Rev. Lett.* **83**, 2498 (1999).
 25. Hall, D. S., Matthews, M. R., Ensher, J. R., Wieman, C. E. & Cornell, E. A. Dynamics of component separation in a binary mixture of Bose-Einstein condensates. *Phys. Rev. Lett.* **81**, 1539–1542 (1998).

Acknowledgements

We thank E. A. Cornell, M. R. Matthews, P. C. Haljan, B. P. Anderson and C. E. Wieman for discussions on the realization of our scheme, and R. Walser and J. Cooper for comments. This work was supported by the National Science Foundation.

Correspondence and requests for materials should be addressed to M.J.H. (e-mail: mholland@hbar.colorado.edu).

Coherent transport of electron spin in a ferromagnetically contacted carbon nanotube

Kazuhiro Tsukagoshi*, Bruce W. Alphenaar† & Hiroki Ago‡

* The Institute of Physical and Chemical Research (RIKEN), Wako, Saitama 351-0198, Japan

† Hitachi Cambridge Laboratory, Madingley Road, Cambridge CB3 0HE, UK

‡ Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

Conventional electronic devices generally utilize only the charge of conduction electrons; however, interest is growing in 'spin-electronic' devices¹, whose operation depends additionally on the electronic spin. Spin-polarized electrons (which occur naturally in ferromagnetic materials) can be injected from a ferromagnet into non-ferromagnetic materials^{2–4}, or through oxide tunnel barriers^{3,5–10}. The electron-scattering rate at any subsequent ferromagnetic/non-ferromagnetic interface depends on the spin polarity, a property that is exploited in spin-electronic devices. The unusual conducting properties^{11–18} of carbon nanotubes offer intriguing possibilities for such devices; their elastic- and phase-scattering lengths are extremely long^{16,17}, and carbon nanotubes can behave as one-dimensional conductors¹⁸. Here we report the injection of spin-polarized electrons from ferromagnetic contacts into multi-walled carbon nanotubes, finding direct evidence for coherent transport of electron spins. We observe a hysteretic magnetoresistance in several nanotubes with a maximum resistance change of 9%, from which we estimate the spin-flip scattering length to be at least 130 nm—an encouraging result for the development of practical nanotube spin-electronic devices.

Figure 1 shows the geometry of our ferromagnetically contacted nanotube devices. We use crude multi-walled carbon nanotubes

(MWNTs) synthesized from graphite rods by the arc discharge evaporation method under a helium atmosphere¹¹. This ensures that the MWNTs contain no trace of magnetic catalysts. The MWNTs are extracted into suspension in dichloroethane by a short sonication, and then dispersed and dried onto a SiO₂/Si substrate. We map out the MWNTs with respect to Pt/Au alignment marks on the substrate using a scanning electron microscope. (The nanotubes show no damage from low-magnification electron microscopy.) The MWNTs are typically 10–40 nm in diameter and 1 µm or more in length. Contact patterns are defined using electron beam lithography, after which a 65-nm-thick layer of Co is deposited by thermal evaporation at a pressure of 4×10^{-7} torr. The resulting polycrystalline Co film behaves ferromagnetically, and has a room-temperature resistivity of approximately $22 \mu\Omega \text{ cm}$. After lift-off, the ferromagnetic leads are connected to non-ferromagnetic Cr/Au bond-pads. Figure 1a is an electron micrograph showing the junction region of a completed device. We have fabricated and characterized more than 50 such devices. Among these, the room-temperature two-terminal resistance varies between 10 and 150 kΩ. When cooled to 4.2 K, the resistance typically increases by a factor of 2–3, in agreement with recently published work¹⁶. The Co contacts remain ohmic up to biases of 10 mV at low temperature.

Magnetoresistance measurements are performed in a 4.2 K bath cryostat with the magnetic field (*B*) from a superconducting magnet directed in the plane of the substrate (*B*_{||}). The two-terminal resistance is measured using an a.c. lock-in technique with an excitation voltage of 100 µV. We note that the lead resistance in our measurements is negligible ($\sim 10 \Omega$) compared with the MWNT resistance and the Co/MWNT contact resistance. The field is swept slowly ($< 10 \text{ mT min}^{-1}$); at this sweep rate the magnetoresistance of test samples contacted by Au/Pt shows no hysteresis in the applied field. Figure 2 shows the two-terminal differential resistance of three different Co-contacted nanotubes as a function of magnetic field. The field is swept first from –100 mT to 100 mT (solid line) and then back to –100 mT (dashed line). In each trace, a resistance peak appears as the magnetic field moves through 0 T. The width of the resistance peak is $\sim 50 \text{ mT}$, which is commensurate with the coercive field strength for a thin Co film¹⁹. There is also a large hysteresis in the peak position ($\sim \pm 50 \text{ mT}$) between positive and negative sweep directions, indicating the probable influence of the contact magnetisation.

Similar hysteretic magnetoresistance is observed in magnetic tunnel junctions (MTJs), where it has been attributed to spin-polarized electron tunnelling^{1,5–10}. MTJs consist of two ferromagnetic contacts separated by a thin oxide layer. The conduction electrons within the ferromagnetic contacts have a preferred spin direction, which is determined by the local magnetization. This causes the formation of majority and minority spin conduction bands with different densities of states at the Fermi energy. In the absence of spin-scattering, the resistance across the tunnel barrier is

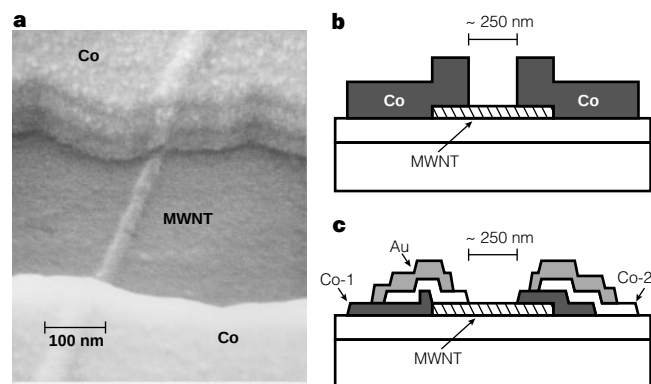


Figure 1 Structure of our devices, consisting of a single multi-walled carbon nanotube (MWNT) electrically contacted by ferromagnetic Co. **a**, Scanning electron microscope image of the device, near the Co/MWNT junction. The Co contacts lie on top of the MWNT, and the conducting channel is 250 nm in length. The image of the nanotube is seen through the Co layer, due to the change in height of the evaporated film. The diameter of this particular nanotube is 30 nm; the diameters varied from 10 to 40 nm. **b**, Schematic diagram of the device. The substrate is a semi-insulating Si wafer covered by a 200-nm-thick layer of SiO₂. The electrodes are polycrystalline Co, deposited by thermal evaporation and defined by electron beam lithography. Non-ferromagnetic leads are positioned $\geq 30 \mu\text{m}$ from the MWNT. The spin polarization of the transport electrons is determined by the local magnetic domains in contact with the MWNT. **c**, Alternative device geometry resulting from a shadow evaporation technique. Note that the Au capping layer does not come in contact with the nanotube, and serves only to protect the ferromagnetic contact layer.

dependent on the relative alignment of the magnetization of the two contacts. In the anti-parallel state the majority spin states are out of alignment and the junction resistance is higher than in the parallel state in which the majority spin states are aligned.

For the nanotube devices in Fig. 2, the contact magnetizations align parallel with the magnetic field at $B = \pm 100$ mT. As we sweep B through 0 T, the magnetization polarity switches. The observed peak in the nanotube resistance suggests that the contact magnetizations switch separately and become misaligned as the field is swept. For a MTJ, misalignment occurs because different ferromagnetic contact materials are used, with different coercivities—the magnetizations are misaligned when B lies between the coercive fields of the two contacts. But this does not explain the misalignment in the nanotube device, because the average coercivity of the two Co contacts should be very similar. In this case, the misalignment may be caused by magnetization fluctuations that occur locally, on the scale of the nanotube diameter (30 nm). The average Co domain size (50 nm; ref. 19) is of the order of the width of the nanotube, so that the nanotube contacts only a small number of magnetic domains. The coercivity of each domain varies, and depends on its geometry and the local energy conditions. Edge and surface effects are also important. It is reasonable, then, that there will be a range of B over which the magnetization at the two nanotube contacts will be misaligned. A resistance peak due to the misaligned state occurs, even though the average properties of the two contacts are similar. The small switches in the resistance, seen most clearly in Fig. 2a, provide additional evidence for local magnetization fluctuations of individual domains.

In Fig. 2 we include data from three different samples to give a fair indication of the large sample-to-sample variation typically observed in the magnetoresistance. In all, 12 of our Co-contacted samples displayed hysteretic magnetoresistance, with a peak height varying from 2% to 10%. Two samples showed a step-like resistance peak as seen in Fig. 2a, while the remaining samples showed a smoother peak as seen in Figs 2b and c. It is likely that the sample-to-sample variations are due to inherent random variations in the surface potential over the small nanotube contact area. Previous experiments on non-magnetically contacted nanotubes have

observed large variations in the contact resistance^{12,16}. Also, in the ferromagnetically contacted samples it is impossible to control the particular domain structure in contact with the nanotube. Reproducibility might be improved by increasing the thickness and quality of the Co layers, which will increase the area of the magnetic domains. Unfortunately, our maximum film thickness (and hence the domain width) is limited to ~ 100 nm by the electron-beam lithographic procedure.

The spin-injection model for the nanotube magnetoresistance requires a sufficiently small amount of spin scattering to occur both within the nanotube, and at the interfaces between the nanotube and the contacts. We can estimate the minimum spin-scattering length in the MWNT using Julliere's model for the MTJ. The difference between the tunnel resistance in the parallel (R_p) and antiparallel (R_a) states⁵ is given by:

$$\Delta R/R_a = (R_a - R_p)/R_a = 2P_1P_2/(1 + P_1P_2)$$

Here, P_1 and P_2 are the percentage of conduction electrons polarized in the majority spin band in the ferromagnetic contacts 1 and 2. For Co, the polarization has been determined⁸ to be 34% giving a maximum resistance change of 21%. In our best case, $\Delta R/R_a$ reaches a maximum value of 9% (Fig. 2b) so that $\sim 14\%$ of the spin-polarized electrons travel the 250 nm through the nanotube without spin-flipping. The spin-scattering length, l_s , can then be estimated by assuming that the spin polarization reduces as $\exp(-l/l_s)$ within the nanotube. This gives $l_s = 130$ nm. Although fairly long, this is probably an underestimation. The spin-polarization near the ferromagnet/nanotube interface will depend on the interface quality, and could be appreciably lower than 34%. Also, we do not take into account spin scattering at the ferromagnet/nanotube interface. Finally, we note that the MTJ theory cannot be expected to completely describe our carbon nanotube device, and a more exact theoretical description is needed.

In an effort to improve the reproducibility of the Co/nanotube contacts, we have fabricated a set of devices using a double evaporation technique (Fig. 1c). Cobalt is evaporated from either side of the substrate to create a continuous Co layer over the nanotube. A gold capping layer is then evaporated from directly

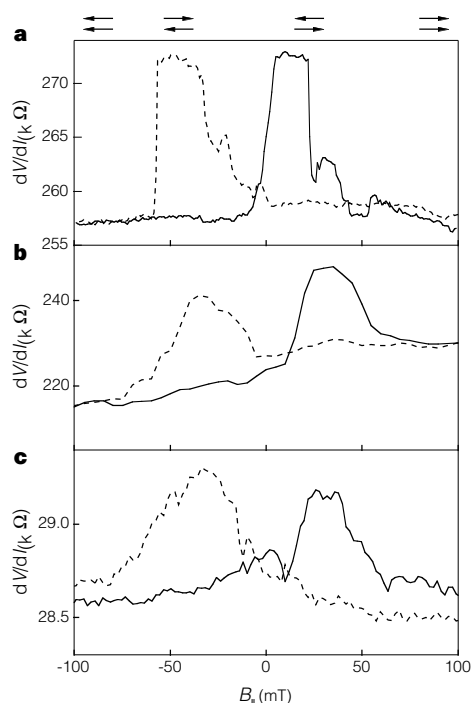
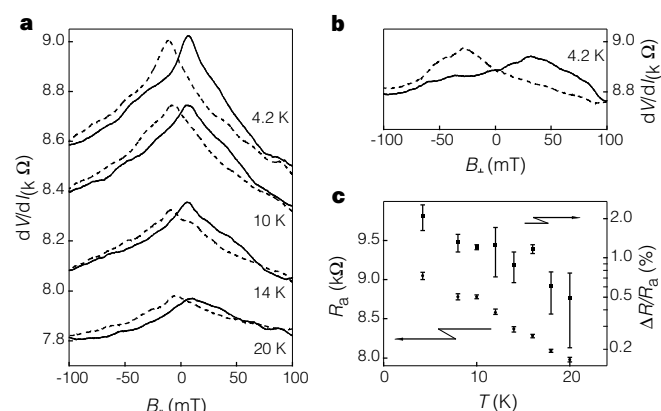


Figure 2 Two-terminal differential resistance as a function of magnetic field for three different MWNT devices. The magnetic field B_s is directed parallel to the substrate, and the temperature is 4.2 K. The solid (dashed) trace corresponds to the positive (negative) sweep direction. The differential resistance shows a large variation among devices—the device shown in **c** has a resistance an order of magnitude lower than the devices shown in **a** and **b**. This is probably due to variations in the contact resistance. The two-terminal resistance could sometimes be lowered by thermal annealing. Each device shows a large hysteretic magnetoresistance peak. We attribute the magnetoresistance peak to spin-polarized injection^{1,5–10} between the ferromagnetic contacts and the MWNT. The magnetization direction of the left and right contacts is represented by the direction of the arrows at the top of the figure. When the magnetizations of the two contacts are parallel, the resistance is lower than when the magnetization of the two contacts are antiparallel. This explanation requires that the spin scattering length in the nanotube is of the order of the contact separation. Also, scattering at the contact interfaces must not completely randomise the spin. The percent difference $\Delta R/R_a$ between the tunnel resistance in the parallel and the antiparallel states is approximately 6% in **a**, 9% in **b** and 2% in **c**.



above the substrate to protect the Co surface. Figure 3a shows the magnetoresistance of a nanotube contacted with a double Co layer as a function of $B_{||}$. This technique improves the continuity of the Co film, and reduces the contact resistance in comparison with the single-Co-layer devices. However, the magnetoresistance ratio $\Delta R/R_a$ and the coercive field are both less than in Fig. 2. Also, rather than observing a sharp magnetoresistance peak, the magnetic field dependence is continuous, similar to what is observed in a granular ferromagnetic film²⁰. This implies that the magnetization is averaged over many small magnetic domains in the double layer Co devices, each with relatively weak coercivities. This averaging improves the stability of the contact, but unfortunately also lowers the magnetoresistance ratio. Figure 3b shows the magnetoresistance as a function of B perpendicular to the substrate ($B_{\perp}$) for the same sample. The peaks are broader and shifted to higher fields when compared with the $B_{||}$ dependence. This is expected for a ferromagnetic film with an in-plane easy axis of magnetization⁶.

The temperature dependence in Fig. 3a shows that the percentage difference between the resistance in the parallel and antiparallel configurations goes to zero as the temperature increases from 4.2 to 20 K. In Fig. 3c we see that $\Delta R/R_a$ decreases approximately exponentially with increasing temperature. The exact mechanism for the temperature dependence is not yet known. Given the low atomic number of carbon, the spin-orbit scattering in the carbon nanotube (and hence its temperature dependence) should be negligible. However, the spin polarization at the interface will decrease with increasing temperature if the nanotube/ferromagnet interface is of relatively poor quality^{8–10}. In addition, the effective contact area will increase with temperature, resulting in the averaging of the spin polarization over an increasing number of magnetic domains. We anticipate that $\Delta R/R_a$ should also decrease with increasing bias across the device, as a similar effect is observed in spin-tunnel junctions⁶. However, our nanotube devices do not survive biases above ~ 10 mV; breakdown occurs before any appreciable change in $\Delta R/R_a$ is observed. Finally, using the double evaporation technique, we have also made a set of devices having two contacts made from different materials, NiFe and Ni. Hysteretic magnetoresistance is observed, with $\Delta R/R_a \approx 2\%$ at 4.2 K.

The results presented here are consistent with recent experiments that show that the electron path lengths within MWNTs are extremely long. The phase coherence length has been found to be 250 nm, and the elastic scattering length 60 nm¹⁶. Further work has shown that MWNTs behave as ballistic conductors, even at room temperature¹⁸. We expect that the spin-flip scattering in the nanotubes should also be highly suppressed. Consider, for example, that the spin scattering length in a metal¹⁴ or a semiconductor²¹ is typically much longer than the phase coherence length or the elastic scattering length. This has allowed progress on a new generation of functional devices whose operation is based on both the charge and

Figure 3 Detailed characterization of a single nanotube device. This device was fabricated using a shadow evaporation technique, resulting in the geometry shown in Fig. 1c. Note that the resistance, magnetoresistance ratio and coercive field are all less than for the devices shown in Fig. 2. **a**, The temperature dependence of the magnetoresistance for magnetic field directed parallel to the substrate ($B_{||}$). The solid (dashed) trace corresponds to the positive (negative) sweep direction. The magnetoresistance ratio $\Delta R/R_a$ decreases approximately exponentially with increasing temperature, but the peak positions remain at a fixed value of $B_{||}$. **b**, The magnetoresistance for field directed perpendicular to the substrate ($B_{\perp}$). The increase in the peak positions and the decrease in the peak heights indicates the presence of an anisotropic domain structure. **c**, The magnetoresistance ratio $\Delta R/R_a$ and R_a for the data set shown in **a**. The error bars take into account the different peak resistance values in the two sweep directions. $\Delta R/R_a$ increases approximately exponentially with decreasing T , while R_a increases approximately linearly.

spin of the electron ('magnetoelectronics')¹. Although we are still some way from producing a practical nanotube spin-electronic device, improvements in the sample preparation should allow the observation of larger and cleaner spin-modulated resistances. Our technique also opens up the possibility of studying magnetically contacted single-walled nanotubes, which could reveal the influence of spin injection on the spin polarization of the carbon nanotube^{22,23}. □

Received 6 May; accepted 2 August 1999.

1. Prinz, G. A. Magnetoelectronics. *Science* **282**, 1660–1663 (1998).
2. Tedrow, P. M. & Meservey, R. Spin-dependent tunnelling into ferromagnetic nickel. *Phys. Rev. Lett.* **26**, 192–195 (1971).
3. Aronov, A. G. Spin injection in metals and polarisation of nuclei. *Pis'ma Zh. Eksp. Teor. Fiz.* **24**, 37–39 (1976); *JETP Lett.* **24**, 32–34 (1976).
4. Johnson, M. & Silsbee, R. H. Interfacial charge-spin coupling—injection and detection of spin. *Phys. Rev. Lett.* **55**, 1790–1793 (1985).
5. Julliere, M. Tunnelling between ferromagnetic films. *Phys. Lett. A* **54**, 225–226 (1975).
6. Moodera, J. S. *et al.* Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. *Phys. Rev. Lett.* **74**, 3273–3276 (1995).
7. Miyazaki, T. & Tezuka, N. Giant magnetic tunnelling effect in Fe/Al₂O₃/Fe junction. *J. Magn. Magn. Mater.* **139**, L231–L234 (1995).
8. Meservey, R. & Tedrow, P. M. Spin-polarised electron tunnelling. *Phys. Rep.* **238**, 173–243 (1994).
9. Moodera, J. S., Nowak, J. & van de Veerdonk, R. J. M. Interface magnetism and spin wave scattering in ferromagnet-insulator-ferromagnet tunnel junctions. *Phys. Rev. Lett.* **80**, 2941–2944 (1998).
10. Shang, C. H., Nowak, J., Jansen, R. & Moodera, J. S. Temperature dependence of magnetoresistance and surface magnetization in ferromagnetic tunnel junctions. *Phys. Rev. B* **58**, R2917–R2920 (1998).
11. Ebbesen, T. W. & Ajayan, P. M. Large-scale synthesis of carbon nanotubes. *Nature* **358**, 220–222 (1992).
12. Langer, L. *et al.* Quantum transport in multiwalled carbon nanotubes. *Phys. Rev. Lett.* **76**, 479–482 (1996).
13. Dai, H., Wong, E. W. & Lieber, C. M. Probing electrical transport in nanomaterials: conductivity of individual carbon nanotubes. *Science* **272**, 523–526 (1996).
14. Ebbesen, T. W. *et al.* Electrical conductivity of individual carbon nanotubes. *Nature* **382**, 54–56 (1996).
15. Saito, R., Dresselhaus, G. & Dresselhaus, M. S. *Physical Properties of Carbon Nanotubes* (Imperial College Press, Singapore, 1998).
16. Bachtold, A. *et al.* in *Proc. 12th Int. Winter School on Electronic Properties of Novel Materials* (eds Kuzmany, H., Fink, J., Mehring, M. & Roth, S.) 65–68 (Am. Inst. Phys., New York, 1998).
17. Bachtold, A. *et al.* Aharonov-Bohm oscillations in carbon nanotubes. *Nature* **397**, 673–675 (1999).
18. Frank, S. *et al.* Carbon nanotube quantum resistors. *Science* **280**, 1744–1746 (1998).
19. Rüdiger, U. *et al.* Magnetoresistance, micromagnetism, and domain-wall scattering in epitaxial hcp Co films. *Phys. Rev. B* **59**, 11914–11918 (1999).
20. Chien, C. L. Magnetism and giant magneto-transport properties in granular solids. *Annu. Rev. Mater. Sci.* **25**, 129–160 (1995).
21. Kikkawa, J. M. & Awschalom, D. D. Resonant spin amplification in n-type GaAs. *Phys. Rev. Lett.* **80**, 4313–4316 (1998).
22. Cobden, D. H. *et al.* Spin splitting and even-odd effects in carbon nanotubes. *Phys. Rev. Lett.* **81**, 681–684 (1998).
23. Tans, S. J. *et al.* Electron-electron correlations in carbon nanotubes. *Nature* **394**, 761–764 (1998).

Acknowledgements

We thank H. Mizuta, T. Nakanishi, M. D. R. Thomas, D. G. Hasko and H. O. Müller for discussions.

Correspondence and requests for materials should be addressed to B.W.A. (alphenaar@phy.cam.ac.uk).

Regional trends in aquatic recovery from acidification in North America and Europe

J. L. Stoddard¹, D. S. Jeffries², A. Lükewille³, T. A. Clair⁴, P. J. Dillon⁵, C. T. Driscoll⁶, M. Forsius⁷, M. Johannessen⁸, J. S. Kahl⁹, J. H. Kellogg¹⁰, A. Kemp¹¹, J. Mannio⁷, D. T. Monteith¹², P. S. Murdoch¹³, S. Patrick¹², A. Rebsdorf¹⁴, B. L. Skjelkvåle⁸, M. P. Stainton¹⁵, T. Traaen⁸, H. van Dam¹⁶, K. E. Webster¹⁷, J. Wieting¹⁸ & A. Wilander¹⁹

¹ Environmental Protection Agency, 200 SW 35th Street, Corvallis, Oregon 97333, USA

² Environment Canada, PO Box 5050, Burlington, Ontario, Canada L7R 4A6

³ Norwegian Institute for Air Research, PO Box 100, 2027 Kjeller, Norway

⁴ Environment Canada, PO Box 6227, Sackville, New Brunswick, Canada E4L 1G6

⁵ Ontario Ministry of the Environment, PO Box 39, Dorset, Ontario, Canada P0A 1E0

⁶ Syracuse University, 220 Hinds Hall, Syracuse, New York 13244, USA

⁷ Finnish Environment Institute, Box 140, 00251 Helsinki, Finland

⁸ Norwegian Institute for Water Research, PO Box 173, 0411 Oslo, Norway

⁹ University of Maine, Sawyer Research Center, Orono, Maine 04469, USA

¹⁰ Vermont Department of Environmental Conservation, 103 S. Main Street, Waterbury, Vermont 05676, USA

¹¹ Environment Canada, 105 McGill, Montreal, Quebec, Canada H2Y 2E7

¹² Environmental Change Research Centre, University College, 26 Bedford Way, London WC1H 0AP, UK

¹³ US Geological Survey, 425 Jordan Road, Troy, New York 12180, USA

¹⁴ National Environmental Research Institute, PO Box 314, 8600 Silkeborg, Denmark

¹⁵ Fisheries and Oceans Canada, 501 University Crescent, Winnipeg, Manitoba, Canada R3T 2N6

¹⁶ Aquasence TEC, PO Box 95125, 1090 Amsterdam, The Netherlands

¹⁷ Wisconsin Department of Natural Resources, 1350 Femrite Drive, Monona, Wisconsin 53716, USA

¹⁸ Umweltbundesamt, Seelstrasse 6–10, 13581 Berlin, Germany

¹⁹ University of Agricultural Sciences, PO Box 7050, 750 07 Uppsala, Sweden

Rates of acidic deposition from the atmosphere ('acid rain') have decreased throughout the 1980s and 1990s across large portions of North America and Europe^{1,2}. Many recent studies have attributed observed reversals in surface-water acidification at national³ and regional⁴ scales to the declining deposition. To test whether emissions regulations have led to widespread recovery in surface-water chemistry, we analysed regional trends between 1980 and 1995 in indicators of acidification (sulphate, nitrate and base-cation concentrations, and measured (Gran) alkalinity) for 205 lakes and streams in eight regions of North America and Europe. Dramatic differences in trend direction and strength for the two decades are apparent. In concordance with general temporal trends in acidic deposition, lake and stream sulphate concentrations decreased in all regions with the exception of Great Britain; all but one of these regions exhibited stronger downward trends in the 1990s than in the 1980s. In contrast, regional declines in lake and stream nitrate concentrations were rare and, when detected, were very small. Recovery in alkalinity, expected wherever strong regional declines in sulphate concentrations have occurred, was observed in all regions of Europe, especially in the 1990s, but in only one region (of five) in North America. We attribute the lack of recovery in three regions (south/central Ontario, the Adirondack/Catskill mountains and midwestern North America) to strong regional declines in base-cation concentrations that exceed the decreases in sulphate concentrations.

In the past two decades, the passage of national (for example, the Clean Air Act in the United States and the Eastern Canada Acid Rain

Program in Canada) and international (for example, the United Nations Economic Commission for Europe's Convention on Long-Range Transboundary Air Pollution (UN/ECE LRTAP) Sulphur Protocols) environmental regulations and agreements has led to widespread declines in the rates of acidic deposition, especially of sulphur, across large regions of North America and Europe. In north/central Europe, SO₂ concentrations in air decreased by 63%, and SO₄²⁻ in precipitation by 40%, between 1985 and 1996 (ref. 2). In the United States and Canada, SO₂ emissions declined 28% between 1980 and 1995 (ref. 1); sulphur deposition in the same period decreased by 29% in the northeastern United States and by 35% in the upper midwestern United States¹. In eastern Canada, decreases in SO₄²⁻ in precipitation ranged from 31% in Atlantic Canada to 36% in south/central Ontario⁵. In Great Britain, both emissions (32%) and wet deposition (43%) of sulphur declined between 1979 and 1993, with most of the decrease observable in the early 1980s (ref. 6). Few of these regions have recorded changes in the rate of nitrogen deposition during this time period. These pervasive declines in rates of acidic deposition create an expectation of widespread recovery in acidified surface waters. Reasoning that consistent recovery across a regional population of streams or lakes would provide the strongest evidence for the success of controls on acidic deposition, we report here on surface water trends at the regional scale.

The primary source of data for this analysis is the International Cooperative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes (ICP-Waters), an international body

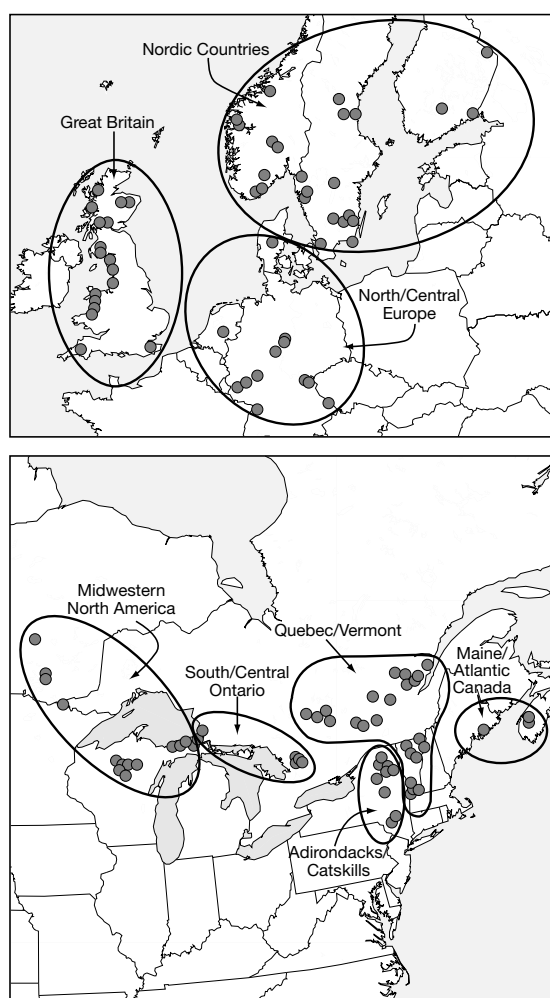


Figure 1 Lake and stream monitoring sites. Data from 205 long-term monitoring sites, located in eight regions of Europe and North America, were used in this analysis.

organized under the UN/ECE LRTAP. Additional data from national, provincial and international monitoring programmes (for example, in the United Kingdom⁷, Canada⁸, the United States⁹ and the International Cooperative Programme on Integrated Monitoring in Europe¹⁰) were added in regions with inadequate sample sizes. The sampling frequencies used by these programmes varied substantially, from weekly in some cases, to quarterly or biannually in others. In order to standardize the data used in trend analyses, we chose quarterly sampling as a reasonable goal (achievable for most sites) and calculated seasonal mean values for the more temporally intensive data sets. We excluded sites with mean Gran alkalinity greater than 200 microequivalents per litre (200 $\mu\text{equiv. l}^{-1}$) in order to focus on those sites most likely to show recovery¹¹, as well as sites with fewer than three quarterly samples (on average), less than 5 years of data per decade, or where climate-driven trends have been documented. A total of 168 sites had sufficient data for the 1980s, and 205 sites had sufficient data for the 1990s. Sites were grouped into regions based on similarities in their geochemistry, in deposition trends in the 1980s and 1990s, and on their geographical proximity (Fig. 1).

Preliminary analyses indicated that temporal patterns in chemical variables were not monotonic; they tended to be curvilinear, with an inflection point around 1990. For this reason, and also to permit a comparison of trends between decades, we performed site-specific trend tests separately on data from the 1980s and 1990s using the non-parametric Seasonal Kendall test modified to account for serial dependence¹². In order to infer regional trends, we applied a meta-analytical technique first described by van Belle and Hughes¹³ that allows, with some restrictions, trend results from multiple sites to be combined into a single estimate of trend.

In performing the regional trend tests, we first tested the Z-statistics from individual Seasonal Kendall tests for homogeneity by attributing the variance among them to each of the possible sources

of interest (sites, seasons and interactions) through an analysis of variance, and tested the resulting sums of squares against a χ^2 distribution¹⁴. There is no accepted threshold for deciding when heterogeneity among sites is sufficient to invalidate combining the trend results regionally. We followed the recommendation of van Belle and Hughes¹³, set the threshold for tests of homogeneity conservatively (for example, $p < 0.01$), and examined the trend statistics more critically if they failed this test. In some cases (particularly for SO_4^{2-}), Z-statistics failed the homogeneity test despite having the same sign (for example, all $Z < 0$). In these cases, we conclude that the evidence for reporting a downward trend is sufficient to regard the results as reliable.

At least two criteria should be fulfilled before extrapolation of site-specific trends to the regional level¹¹: (1) the sites must be representative of the region or the regional subpopulation (for example, acid-sensitive surface waters) of interest; and (2) the sites should exhibit consistency of trend behaviour. We can test for the latter condition through the homogeneity test described above, but we have no current test for the former condition. Here we assume that monitoring programmes have made good judgments in choosing sites representative of acid-sensitive surface waters in each region. All of the ICP-Waters sites were selected using published criteria (for example, sites in headwater regions, free from local disturbance, and located on sensitive geology)¹⁵. The national and provincial programmes presented in this analysis used similar criteria.

The strongest evidence for a regional surface-water response to decreasing deposition comes from SO_4^{2-} trends (Table 1). Sulphate (we used non-marine SO_4^{2-} in regions within 100 km of oceans) has declined across most of Europe and North America, generally with more rapid declines during the 1990s. Only Great Britain failed to show a significant decline in SO_4^{2-} in the 1990s. Further, SO_4^{2-} trends were generally homogeneous within a region, suggesting strongly that SO_4^{2-} trends are driven by changes in atmospheric deposition. Only two subsets of data failed the homogeneity test. For south/central Ontario during the 1980s, all of the Z-statistics were negative, indicating decreasing SO_4^{2-} , and we consider the regional trend to be valid despite high χ^2 values. Vermont/Quebec in the 1980s was the only region that exhibited heterogeneity strong enough to invalidate a significant trend test; all of the sites with increasing SO_4^{2-} in this region and decade were located in Quebec, although none of the positive site-specific trends were significant.

Nitrate increases were nearly universal, but largely restricted to the 1980s. The largest increases, and probably the only ecologically significant ones, occurred in the Adirondack/Catskill mountains and north/central Europe. The nitrate increases in these regions during the 1980s have received considerable attention, generating concern that these high-deposition regions are experiencing nitrogen saturation¹⁶. Interestingly, these two regions also show the largest reversals of NO_3^- trends in the 1990s.

Both conceptual¹⁷ and mechanistic¹⁸ models of acidification suggest that decreases in acid anion concentrations ($\Sigma[\text{SO}_4^{2-} + \text{NO}_3^-]$) should be balanced by smaller decreases (relative to acid anions) in base cations (C_B) and increases in alkalinity (which we define as recovery). Large-scale patterns in recovery were evident across all the European regions (north/central Europe, the Nordic countries and Great Britain) during the 1990s (Table 1, Fig. 2). The lack of recovery in the Nordic countries in the 1980s has been reported elsewhere¹⁹, and declines in C_B identified as the likely mechanism preventing alkalinity from recovery despite large decreases in sulphur deposition. Our analysis (Fig. 2) reinforces this interpretation, but highlights how this pattern changed in the 1990s. Nordic trends in C_B in the 1990s are relatively flat (Table 1), and continued strong decreases in SO_4^{2-} in this region appear to be leading to a recovery in alkalinity. It should be noted that the "recovery" observed in Great Britain does not appear to be driven by changes in acid anions (non-marine SO_4^{2-} is unchanged in the

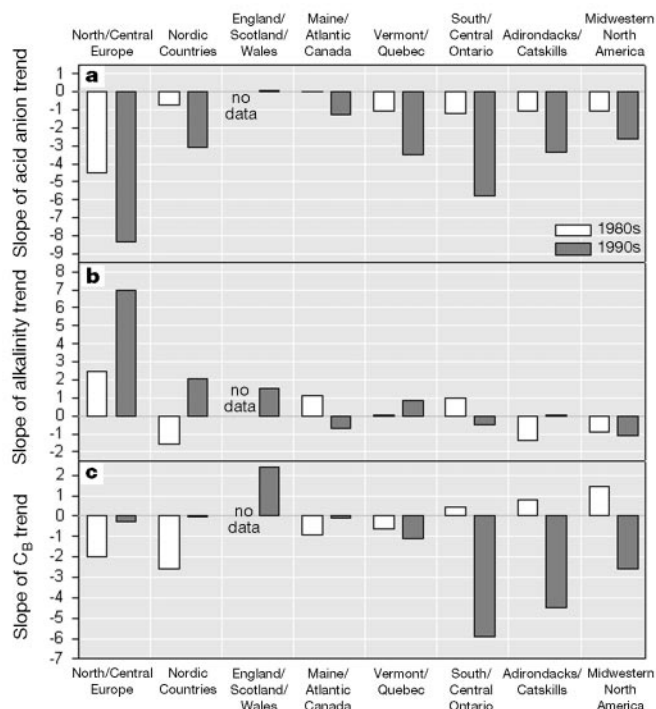


Figure 2 Regional acidification trend results. Histograms exhibit slopes of regional trends in **a**, acid anions ($\text{SO}_4^{2-} + \text{NO}_3^-$), **b**, measured (Gran) alkalinity, and **c**, base cation (C_B) concentrations for eight regions of Europe and North America for the 1980s and 1990s. All slopes are in $\mu\text{equiv. l}^{-1} \text{yr}^{-1}$.

Table 1 Regional trend

Region	Decade	N*	SO ₄ ²⁻ trend		NO ₃ trend		Alkalinity trend		C _B trend	
			<i>p</i>	Slope (μequiv. l ⁻¹ yr ⁻¹)	<i>p</i>	Slope (μequiv. l ⁻¹ yr ⁻¹)	<i>p</i>	Slope (μequiv. l ⁻¹ yr ⁻¹)	<i>p</i>	Slope (μequiv. l ⁻¹ yr ⁻¹)
North/central Europe	1980s	12	<0.001	-5.8	0.011	+1.3	0.089	+2.5	0.921	-2.0
	1990s	23	<0.001	-5.9	<0.001	-2.4	<0.001	+7.0	0.909	-0.3
Nordic countries	1980s	27	<0.001	-0.8	0.002	+0.1	<0.001	-1.6†	<0.001	-2.6‡
	1990s	33	<0.001	-3.1	0.418	-0.0	<0.001	+2.1	0.778	-0.0
Great Britain	1990s	18	0.510	+0.0	0.024	+0.1	<0.001	+1.5	<0.001	+2.4
Maine/Atlantic Canada	1980s	10	0.264	-0.0	0.829	-0.0	<0.001	+1.2	0.180	-0.9
	1990s	10	<0.001	-1.3	0.883	-0.0	0.111	-0.7	0.685	-0.1
Vermont/Quebec	1980s	47	<0.001	-1.1†	<0.001	+0.0	0.515	+0.1	<0.001	-0.6
	1990s	49	<0.001	-3.5	0.133	-0.0	0.028	+0.9	0.007	-1.1
South/central Ontario	1980s	13	<0.001	-1.2‡	0.002	+0.0§	<0.001	+1.0†	0.030	+0.4
	1990s	13	<0.001	-5.8	0.591	-0.0	0.817	-0.5	<0.001	-5.9‡
Adirondack/Catskill mountains	1980s	25	<0.001	-1.7	<0.001	+0.7	<0.001	-1.3	0.775	+0.8
	1990s	25	0.012	-0.9	<0.001	-2.4	0.426	+0.1	<0.001	-4.5
Midwestern North America	1980s	34	0.023	-1.2	<0.001	+0.1§	0.123	-0.9	<0.001	+1.5
	1990s	34	<0.001	-2.6	0.782	-0.0	<0.001	-1.1	<0.001	-2.8

Results of meta-analysis of trends in SO₄²⁻, NO₃, alkalinity and base cation (C_B) concentration for 1980s and 1990s in various regions. Statistics (*p* values based on χ^2 tests) are from an analysis of variance of site trend *Z* scores, and constitute a test of trend homogeneity within the regions. χ^2 values were considered significant at *p* < 0.01 for site and season effects. Statistics for a site–season interaction term are not shown, as they were never significant. Significant trends (*p* < 0.05) are shown in bold.

* Sample sizes vary among variables. *N* values listed are for SO₄²⁻.

† Site heterogeneity is significant; trend statistics are questionable.

‡ Site heterogeneity is significant, but all slopes are negative; trend statistics considered valid.

§ Seasonal heterogeneity significant; trend statistics questionable.

|| Seasonal heterogeneity significant, but all slopes are positive; trend statistics considered valid.

1990s, while NO₃ has increased slightly), but instead results from increasing C_B concentrations (this is the only region to exhibit this phenomenon). The pattern in Great Britain appears to be driven by climate variations, particularly as they affect the transport of seasalt aerosols across the region, rather than by changes in anthropogenic emissions²⁰.

Despite large decreases in acid anions in south/central Ontario, the Adirondack/Catskill mountains and midwestern North America in the 1990s, there was either no regional recovery in alkalinity or continued acidification (Fig. 2); we attribute this pattern to compensating negative trends in C_B. Although some decrease in C_B is expected as SO₄²⁻ concentrations decline, C_B declines in excess of SO₄²⁻ declines are unexpected, and will preclude any recovery in alkalinity²¹. Among the mechanisms that have been hypothesized to produce excessive C_B declines are (1) decreasing C_B deposition²², and (2) cation depletion of watershed soils^{23,24}. While the former mechanism has been discounted in at least some regions¹⁴, it is possible that long-term high rates of acidic deposition have leached enough cations from sensitive soils that adsorbed cation pools are severely depleted. If this depletion continues even at diminished levels of acidic deposition, it could produce the steep C_B trends we observe in these regions. The possibility that this mechanism will prevent or delay recovery deserves critical attention.

It is important to consider the effects that climate variation could have on the trends we report, especially for the relatively short 1990s time series we use in our analysis. The effects of climate on non-marine SO₄²⁻ trends and C_B behaviour in Great Britain have already been mentioned. Droughts in both south/central Ontario and midwestern North America have been shown to cause re-oxidation of reduced sulphur in wetlands and lake littoral regions^{25,26}. The resulting increase in sulphate export can be relatively long-term, temporarily delaying the recovery response in lakes with plentiful wetlands in their watersheds; this mechanism undoubtedly contributes to the lack of homogeneity in SO₄²⁻ trends observed in south/central Ontario. In midwestern North America, however, where long water residence times create a great potential for climatic effects²⁷, declines in SO₄²⁻ in the 1990s far exceed what is attributable to a recovery from drought in the 1980s²⁸; these trends doubtless reflect declining rates of deposition. Climate may also play a role in

producing strongly decreasing C_B trends. In lakes with long residence times, prolonged periods of drought produce increased C_B concentrations and decreased rates of C_B export²⁹; as lakes recover from drought, C_B concentrations may decline rapidly (for example, in drainage lakes, where increases result primarily from evapotranspiration) or slowly (for example, in seepage lakes, where increases result from changes in groundwater flowpaths)²⁷. Although data from the current study are not sufficient to identify mechanisms incontrovertibly, the widespread declines in SO₄²⁻ we report are completely consistent with observed changes in rates of sulphur deposition in Europe and North America. Lack of lake and stream recovery, and particularly the strong decreases in C_B concentrations observed in some regions, may be due to a number of mechanisms, including climate fluctuations.

The three North American regions where recovery might be expected but was not observed (south/central Ontario, Adirondack/Catskill mountains and midwestern North America) exhibit the largest North American declines in acid anions and largest decreases in C_B in the 1990s, and are closest to significant emissions sources in the midwestern United States. Analysing whether the proximity of these regions to significant sources (and therefore higher, and possibly longer-term, inputs of acidic deposition) is related to their C_B behaviour is beyond the scope of this Letter. These trend patterns are very similar to those observed in the 1980s in the Nordic countries, where recovery is now occurring. Increasing trends in alkalinity and unchanging concentrations of C_B in the Nordic countries during the 1990s may be the result of higher rates of SO₄²⁻ decline (Fig. 2a), or may reflect the cumulative effects of longer-term declines in sulphur deposition (beginning in the late 1960s) in this region³⁰. Long-term decreases in sulphur deposition could be expected to result in the recovery of soil cation pools (as rates of primary weathering begin to exceed loss rates due to leaching by acid anions), and eventual recovery after a sufficient time lag¹⁷. The recovery pattern in the Nordic countries suggests that larger decreases in sulphur deposition and/or a longer response time may be required before similar recovery is widely observed in North America. This suggestion of lagged recovery, coupled with evidence of the deleterious effects of climate variability in detecting recovery, highlights the importance of continued coordinated

international monitoring to assess the success of acidic deposition control measures. □

Received 20 April; accepted 19 August 1999.

- Shannon, J. D. Regional trends in wet deposition of sulfate in the United States and SO₂ emissions from 1980 through 1995. *Atmos. Environ.* **33**, 807–816 (1999).
- Tarrasson, L. (ed.) *Transboundary Acidifying Air Pollution in Europe* (International Cooperative Programme for the Monitoring and Evaluation of Long Range Transmission of Air Pollutants in Europe/Meteorological Synthesizing Centre-West, 1998).
- Skjelkvåle, B. L., Wright, R. F. & Henriksen, A. Norwegian lakes show widespread recovery from acidification; results from national surveys of lakewater chemistry 1986–1997. *Hydrol. Earth System Sci.* **2**, 375–577 (1998).
- Driscoll, C. T., Likens, G. E. & Church, M. R. Recovery of surface waters in the northeastern U.S. from decreases in atmospheric deposition of sulfur. *Wat. Air Soil Pollut.* **105**, 319–329 (1998).
- Summers, P. W. Time trend of wet deposition acidifying potential at five ecological monitoring sites in Eastern Canada 1981–1993. *Wat. Air Soil Pollut.* **85**, 653–658 (1995).
- Downing, C. E. H., Vincent, K. J., Campbell, G. W., Fowler, D. & Smith, R. I. Trends in wet and dry deposition of sulphur in the United Kingdom. *Wat. Air Soil Pollut.* **85**, 659–664 (1995).
- Patrick, S., Monteith, D. T. & Jenkins, A. *UK Acid Waters Monitoring Network: The First Five Years* (Department of the Environment, ENSIS Publishing, London, 1995).
- Jeffries, D. S. 1997 *Canadian Acid Rain Assessment Vol. 3, Aquatic Effects* (Environment Canada, Ottawa, 1997).
- Ford, J., Stoddard, J. L. & Powers, C. F. Perspectives in environmental monitoring: an introduction to the U.S. EPA Long-Term Monitoring (LTM) project. *Wat. Air Soil Pollut.* **67**, 247–255 (1993).
- Kleemola, S. & Söderman, G. *Manual for Integrated Monitoring, Programme Phase 1993–1996* (Environmental Data Center, National Board of Waters and the Environment, Helsinki, 1993).
- Stoddard, J. L., Driscoll, C. T., Kahl, J. S. & Kellogg, J. Can site-specific trends be extrapolated to a region? An acidification example for the Northeast. *Ecol. Appl.* **8**, 288–299 (1998).
- Hirsch, R. M. & Slack, J. R. A nonparametric trend test for seasonal data with serial dependence. *Wat. Resour. Res.* **20**, 727–732 (1984).
- van Belle, G. & Hughes, J. P. Nonparametric tests for trend in water quality. *Wat. Resour. Res.* **20**, 127–136 (1984).
- Stoddard, J. L., Driscoll, C. T., Kahl, J. S. & Kellogg, J. A regional analysis of lake acidification trends for the northeastern U.S., 1982–1994. *Environ. Monit. Assess.* **51**, 399–413 (1998).
- International Cooperative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes. *Programme Manual* (Programme Centre, Norwegian Inst. for Water Research, Oslo, 1995).
- Stoddard, J. L. in *Environmental Chemistry of Lakes and Reservoirs* (ed. Baker, L. A.) 223–284 (American Chemical Society, Washington DC, 1994).
- Galloway, J. N., Norton, S. A. & Church, M. R. Freshwater acidification from atmospheric deposition of sulfuric acid: A conceptual model. *Environ. Sci. Technol.* **17**, 541–545 (1983).
- Cosby, B. J., Hornberger, G. M., Galloway, J. N. & Wright, R. F. Time scales of catchment acidification: A quantitative model for estimating freshwater acidification. *Environ. Sci. Technol.* **19**, 1144–1149 (1985).
- Kirchner, J. W. & Lydersen, E. Base cation depletion and potential long-term acidification of Norwegian catchments. *Environ. Sci. Technol.* **29**, 1953–1960 (1995).
- Evans, C., Monteith, D. T. & Harriman, R. Long-term variability in the deposition of marine ions at west coast sites in the U.K. Acid Waters Monitoring Network: Impacts on surface water chemistry and significance for trend determination. *Sci. Tot. Environ.* (in the press).
- Driscoll, C. T., Likens, G. E., Hedin, L. O., Eaton, J. S. & Bormann, F. H. Changes in the chemistry of surface waters: 25-year results at the Hubbard Brook Experimental Forest. *Environ. Sci. Technol.* **23**, 137–143 (1989).
- Hedin, L. O. et al. Steep declines in atmospheric base cations in regions of Europe and North America. *Nature* **367**, 351–354 (1994).
- Likens, G. E., Driscoll, C. T. & Buso, D. C. Long-term effects of acid rain: Response and recovery of a forest ecosystem. *Science* **272**, 244–246 (1996).
- Lawrence, G. B. et al. Soil calcium status and the response of stream chemistry to changing acidic deposition rates in the Catskill Mountains, New York. *Ecol. Appl.* (in the press).
- Bayley, S. E., Schindler, D. W., Parker, B. R., Stainton, M. P. & Beaty, K. G. Effects of forest fire and drought on the acidity of a base-poor boreal forest stream: similarities between climatic warming and acidic precipitation. *Biogeochemistry* **17**, 191–204 (1992).
- Dillon, P. J., Molot, L. A. & Futter, M. A note on the effects of El Niño-related drought on the recovery of acidified lakes. *Int. J. Environ. Monit. Assess.* **46**, 105–111 (1997).
- Webster, K. E. et al. Structural features of lake districts: Landscape controls on lake chemical responses to drought. *Freshwat. Biol.* (in the press).
- Webster, K. E. & Brezonik, P. O. Climate confounds detection of chemical trends related to acid deposition in Upper Midwest lakes in the USA. *Wat. Air Soil Pollut.* **85**, 1575–1580 (1995).
- Schindler, D. W. et al. The effects of climatic warming on the properties of boreal lakes and streams at the Experimental Lakes Area, northwestern Ontario. *Limnol. Oceanogr.* **41**, 1004–1017 (1996).
- Mylona, S. *Trends of Sulphur Dioxide Emissions, Air Concentrations and Depositions of Sulphur in Europe since 1880* (International Cooperative Programme for the Monitoring and Evaluation of Long Range Transmission of Air Pollutants in Europe/Meteorological Synthesizing Centre-West, 1994).

Acknowledgements

We thank the LRTAP Working Group on Effects for their support; this Working Group supports the production of international, quality-controlled, comparable data. We also acknowledge the work of the ICP Programme Centre at the Norwegian Institute of Water Research (NIVA), where the data are collated, verified and archived, and thank T. J. Sullivan and M. R. Church for comments and suggestions. This work was supported by the US Environmental Protection Agency.

Correspondence and requests for materials should be addressed to J.L.S. (e-mail: stoddard@mail.cor.epa.gov).

Sterilization and canopy modification of a swollen thorn acacia tree by a plant-ant

Maureen L. Stanton^{*†}, Todd M. Palmer^{†‡}, Truman P. Young^{*†‡}, Amanda Evans[§] & Monica L. Turner[†]

^{*} Center for Population Biology, University of California, Davis, California 95616, USA

[†] Mpala Research Centre, PO Box 555, Nanyuki, Kenya

[‡] Department of Environmental Horticulture, University of California, Davis, California 95695, USA

[§] Department of Biology, University of Oregon, Eugene, Oregon 97403, USA

Obligate symbioses between specialized arboreal ants and plants have evolved independently in many lineages^{1,2}. Ant-plants (myrmecophytes) typically provide hollow nest cavities and nutrition to the occupying ant colony^{1,3–6}. In turn, resident plant-ants often protect their hosts from herbivory^{7–11} and/or overgrowth by surrounding vegetation^{12,13}. As individual plants are rarely occupied by more than one ant colony^{14–17}, co-occurring plant-ant species compete intensely for hosts^{13,14,18,19}. In such multi-species systems, ecological interactions among potential partners may lead to the evolution of cheating^{20,21}. Previous studies have revealed that some specialized plant-ants are effectively parasites of their host-plants^{8,18,22,23}, but the selection pressures favouring such behaviours are poorly understood. Here we describe host parasitism in an east African plant-ant that prunes and sterilizes its host-tree canopies, apparently to minimize contact with competitively dominant ants occupying neighbouring trees. We propose that the high density of ant-trees and low diversity of tree species in this savanna habitat have selected for induced, parasitic pruning of host trees by this competitively subordinate ant species.

Whistling Thorn (*Acacia drepanolobium*) trees dominate vast areas of savanna on black cotton soils in upland east Africa^{24,25}. At each branch node, *A. drepanolobium* trees produce either a pair of slender thorns or a swollen thorn pair joined by a bulbous, hollow base 1.5–6 cm in diameter (Fig. 1). When living on a tree, ants chew



Figure 1 Close-up of *A. drepanolobium* branches showing axillary leaves, paired slender thorns and a node occupied by a fused swollen thorn pair. To show architectural features clearly, this photograph was taken at the beginning of the rainy season, when relatively few axillary leaves were persistent on older growth. Resident ants had not yet chewed entry holes into the recently formed swollen thorn (centre). Swollen thorns are produced even in the absence of ants²⁶.

entry holes into the swollen thorn bases and use the hollow interiors to house workers and raise brood^{25,26}. Four species of specialized plant-ants co-exist within most populations of *A. drepanolobium*²⁶. In our study site at the Mpala Research Centre in Kenya (36° 50' E, 0° 15' N), *Crematogaster mimosae* occupies almost 50% of the trees, while the remaining trees are divided between three other ant species: *Crematogaster sjostedti*, *Crematogaster nigriceps*, and *Tetraponera penzigi*. Less than 1% of trees over 0.5 m tall are unoccupied. Three of these ant species occur only on *A. drepanolobium* trees within this ecosystem; *C. sjostedti* also occupies *Acacia seyal*, a far less abundant tree which can also produce swollen thorns²⁵.

Canopy architecture varies significantly among *A. drepanolobium* trees occupied by the four ant species in our study site (Wilk's Lambda: $P = 0.0179$, $F = 3.174$; degrees of freedom (d.f.) = 15, 14). Trees occupied by *C. nigriceps* are characterized by an unusually high degree of branching, reduced apical dominance and lateral spread, large swollen thorns and high swollen thorn density (Fig. 2). There are two alternative hypotheses that could explain this pattern. *C. nigriceps* colonies may selectively occupy trees with this type of architecture; or *C. nigriceps* may modify the architecture of host-tree canopies to increase secondary branching at the expense of lateral canopy spread.

Field observations provide evidence for active canopy modification by *C. nigriceps*. *A. drepanolobium* trees produce dense 'cushions'²⁴ of axillary buds below each pair of stipular thorns. Cushions remain active for at least two years, producing leaves, branches and/or flowers during subsequent growth cycles. At our study site, trees occupied by *C. nigriceps* are uniquely characterized by the absence of most axillary cushions²⁵. Undamaged cushions occur principally at nodes with swollen thorns, and it is from these nodes that most leaves and lateral branches arise. We observed that *C. nigriceps* workers aggregate on growing shoots of the host tree, where they rapidly destroy most axillary cushions and eventually kill the apical meristem. These pruning activities probably cause the unique canopy architecture associated with *C. nigriceps*. Moreover, pruning sterilizes the host-tree; destruction of cushions and occupancy by *C. nigriceps* are associated with greatly reduced flower production by *A. drepanolobium*²⁵.

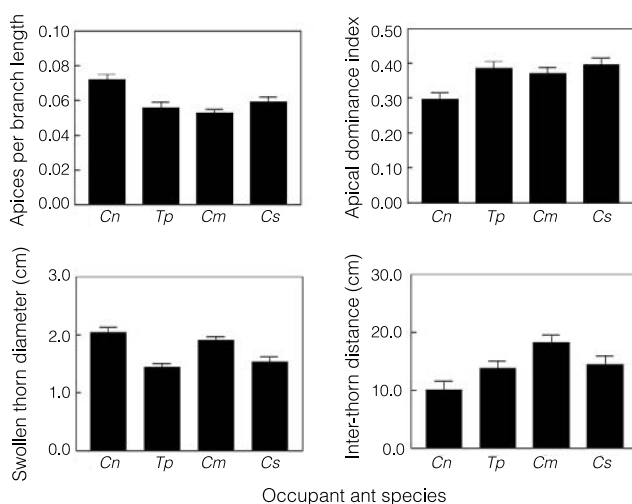


Figure 2 Canopy attributes of *A. drepanolobium* trees occupied by four different species of acacia ants. Ant species: Cn, *C. nigriceps*; Tp, *T. penzigi*; Cm, *C. mimosae*; Cs, *C. sjostedti*. For five architectural measurements, multivariate analysis of variance showed that canopy architecture varies significantly among occupant ant species; least-squares means and standard errors shown are based on that analysis. In this study, the length of slender thorns did not vary significantly among species of ant occupants and so is not shown.

To test the hypothesis that *C. nigriceps* ants actively modify the branching architecture of their host trees, we assigned one-half of the trees used for architectural measurements to an ant-removal treatment, while the other half served as controls. Two swollen thorn nodes on recent growth were selected at random on each tree. The number of branches emerging from each node were counted before treatments in June 1996, and after two full growing seasons in December 1997.

The ant-removal experiment confirmed that *C. nigriceps* significantly increase branching in canopies of their host trees. On control trees continuously occupied by *C. nigriceps*, branch production at swollen thorn nodes increased over 40% from 1996 to 1997. In contrast, on trees from which *C. nigriceps* had been removed, the numbers of branches per node decreased 25% over the same interval. These treatment-dependent changes in architecture were virtually absent on trees occupied by *C. sjostedti*, *C. mimosae* or *T. penzigi*. Most axillary bud cushions were destroyed on the post-1996 growth of control trees occupied continuously by *C. nigriceps*, but not on other trees. None of these trees flowered during the course of the experiment.

Competition for host trees leads to violent conflicts between the four species of acacia ant occupying *A. drepanolobium*. At our study site, we have observed a number of take-over raids in which ants from nearby *C. mimosae* or *C. sjostedti* colonies stream onto trees occupied by either *C. nigriceps* or *T. penzigi* and attempt to dislodge workers and brood from inside the swollen thorns. Field experiments confirmed that the four ant species vary markedly in their ability to take over trees forcefully and to defend their trees against take-overs. We staged conflicts between colonies of different species by tying together branches of two adjacent trees, each occupied by a different ant species. *C. nigriceps* was evicted from its host tree more than any other ant, losing 80% of the time in conflicts against *C. sjostedti* and 64% of the time against *C. mimosae* (Fig. 3). In contrast, although *T. penzigi* was rarely the aggressor, it successfully defended its host tree from attack 71% of the time. These experiments show that *C. nigriceps* colonies are unusually vulnerable to take-over when their host trees come into contact with colonies of the competitively dominant ants *C. sjostedti* and *C. mimosae*.

As trees grow taller and their canopies spread laterally, contacts

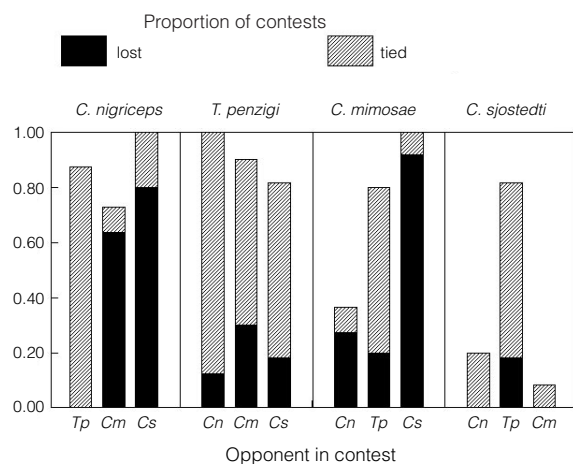


Figure 3 Outcomes of staged conflicts between colonies of different ant species inhabiting neighbouring *A. drepanolobium* trees. There were 8–12 encounters between each pair of ant species. The species listed across the top of the Figure was the loser if it was displaced by the opposing colony from most or all of its joined host tree. A tie was scored if both colonies retained their host trees during six months following joining. Species of opposing colonies (below x axis): Cn, *C. nigriceps*; Tp, *T. penzigi*; Cm, *C. mimosae*; Cs, *C. sjostedti*.

with neighbouring tree canopies and potentially hostile ant colonies become more likely. In a random survey of 1,796 trees throughout the study site, we found that the probability of canopy contact with adjacent trees increases from 6.3% in trees less than 1.0 m to 34.0% in trees more than 3 m in height. Tall trees are predominately occupied by *C. sjostedti* and *C. mimosae*²⁵, so the lateral canopy spread associated with tree growth presents a significant danger to competitively subordinate colonies of *C. nigricaps*.

As trees occupied by *C. nigricaps* grow, they show significantly less lateral canopy spread than trees occupied by the other three ant species. Analysis of covariance showed that canopy spread increases significantly with height ($P < 0.0001$; $F = 378.75$, d.f. = 1,294), and varies significantly among ant occupants ($P = 0.0056$; $F = 4.29$, d.f. = 3,294), with *C. nigricaps* trees showing the least lateral spread. The significant interaction between height and ant occupant ($P < 0.0001$; $F = 11.28$; d.f. = 3,294) revealed that the increase in lateral spread with height in trees occupied by *C. nigricaps* (0.296) is roughly half that for trees occupied by other ants (0.532–0.72). Quadratic regressions showed that trees occupied by *C. nigricaps* have diminishing rates of lateral canopy growth with increasing height, whereas canopy spread and height are linearly related in trees occupied by the other ant species. Together, these observations and experiments indicate that canopy pruning by *C. nigricaps* colonies may reduce potential contacts between their host trees and ant colonies on neighbouring trees.

In the *A. drepanolobium* system, detection of neighbouring ant colonies should be possible because at least three of the ant species produce potent alarm pheromones when distributed^{26,27}. Moreover, because shoot pruning is time consuming for the resident ant colony and severely detrimental to the host tree, one might expect pruning to be conditionally induced by the presence of potentially aggressive colonies nearby. To test this hypothesis, we surveyed a random sample of 555 trees to determine whether canopy geometry is affected by proximity to neighbouring trees and ant colonies. As predicted, canopies of *C. nigricaps* trees were highly asymmetrical, showing significantly less lateral spread in the direction of neighbouring trees occupied by different ant colonies, and tending to extend outwards into open space or towards nearby trees occupied by the same colony ($P = 0.0047$; Fig. 4). In contrast, the canopy geometry of trees occupied by the other three ant species was independent of neighbouring trees and colonies ($P > 0.45$ in all

cases). These measurements indicate that *C. nigricaps* may selectively prune their host trees to avoid canopy contact with neighbouring ant colonies.

We propose that the evolution of host-tree pruning and sterilization by *C. nigricaps* ants has been driven by three attributes of the black cotton savanna community: the high density of ant-acacia trees; low tree-species diversity; and the presence of competitively dominant plant-ants using the same host-tree species. Other studies have shown that competitively subordinate plant-ants prune surrounding vegetation away from their hosts, presumably to eliminate arboreal pathways by which dominant ants can attack¹⁹. Because this type of pruning behaviour also reduces the susceptibility of the host plant to fire and overgrowth^{12,28}, it poses no threat to the plant-ant mutualism. In contrast, because the canopy in black cotton habitats is virtually a monoculture of *A. drepanolobium* trees, workers would certainly be attacked by aggressive, resident colonies if they attempted to prune the canopies of nearby trees. Instead, competitively subordinate colonies of *C. nigricaps* prune and consequently sterilize their own host tree to minimize the risk of contact with more dominant ants. Our study reinforces the view that there is a fine line between mutualism and parasitism^{20,21}, and that the attributes of the ecological community as a whole may determine the direction in which any given ant-plant association evolves¹³. □

Methods

Analysis of architectural traits

For each of 242 trees 0.75–2.0-m tall along four 200-m transects, we measured the base diameters of swollen thorns, lengths of slender thorns and branch distances between recently produced swollen thorns at two randomly canopy locations. On two branch systems per tree, we counted apices per total branch length to estimate the amount of branching, and measured the tendency for new growth to spread laterally away from the main stem as the ratio of longest shoot length to summed secondary shoot lengths (the 'apical dominance index'). The General Linear Models procedure of SAS²⁹ was used to conduct a multivariate analysis of variance (ANOVA); the effect of ant species (a fixed factor) was tested over the ant species-by-transect interaction. The effects of transect and ant-by-transect interaction were not statistically significant ($P > 0.30$).

Ant-removal methods

For trees assigned to the ant-removal treatment, all swollen thorns were cut open. Ants were brushed from the canopy, and a barrier of Tanglefoot and Teflon tape was placed around the tree's stem to prevent recolonization. On control trees we cut off the pair of slender thorns nearest to each swollen thorn; ants were distributed but not removed.

Statistical analysis of the effects of ant removal on branching

As there was no significant variation in branching between trees occupied by *T. penzigi*, *C. mimosae*, and *C. sjostedti*, these three species were pooled in a repeated-measures ANOVA²⁹. Occupant ant species (*C. nigricaps* versus others) and treatment (ant removal versus control) were the between-subjects factors in the model; year was the within-subject factor. The number of branches per swollen thorn node was much greater in trees originally occupied by *C. nigricaps* than those occupied by the other three species ($P < 0.0001$; $F = 62.75$; d.f. = 1,151). The significant year-by-treatment interaction ($P < 0.002$) showed that the ant-removal treatment increased branching over time, principally owing to trees occupied by *C. nigricaps*. In the overall ANOVA, there was a significant interaction between year, initial *C. nigricaps* occupancy and ant-removal treatment ($P = 0.0104$; $F = 6.729$; d.f. = 1,151). In a second ANOVA conducted just for the other three ant species, the interaction between year and removal treatment was not significant ($P = 0.321$; $F = 0.993$; d.f. = 2,110).

Staging of inter-colony conflicts

In staging conflicts between neighbouring colonies of different species, we matched experimentally joined host trees by height and numbers of swollen thorns. Each pair of ant species was used in 8–12 trials. Colonies of the *Crematogaster* species often occupy multiple trees, and we did not prevent recruitment from other trees during the experiment. A tie was scored for each participating species when neither colony displaced the other from the majority of its joined tree over 6 months. Take-over of a joined tree resulted in a win and a loss.

Analysis of lateral canopy spread and tree height

Stratified sampling was necessary to achieve comparable sample sizes for each ant species across tree height classes. Along random transects, we selected one of four target height classes (1–1.99 m, 2–2.99 m, 3–3.99 m, ≥ 4 m) at random, and then searched until we found the nearest tree within that height class occupied by each of the plant-ant species.

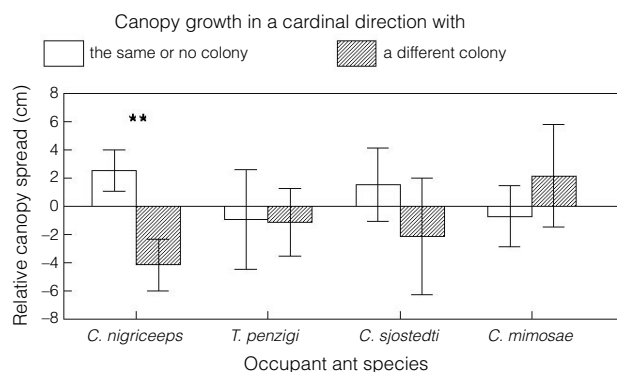


Figure 4 Growth asymmetry of *A. drepanolobium* canopies, as a function of ant occupant and the presence of hostile ant colonies on neighbouring trees. A positive value of relative canopy spread indicates disproportionate lateral growth in a given direction. The local neighbourhood within each quadrant centred around a focal tree was categorized as 'different', if a tree bearing a different colony occurred within 3 m; or 'same or none', if no different ant colony occurred within 3 m. Separate ANOVAs were conducted for each ant species; least-square means and standard errors from those models are shown. Double asterisks, $P < 0.01$.

The height of the tallest living shoot was measured, and canopy diameter was calculated as the mean of two perpendicular canopy widths. Analysis of covariance was used²⁷ to determine whether mean canopy spread is predicted by tree height and/or species of ant occupant. For each ant, canopy spread was regressed on tree height and the square of height. The quadratic component was significantly negative for trees occupied by *C. nigriceps* ($P = 0.0307$), but not for trees occupied by the other three ant species (P ranged from 0.16 to 0.84).

Analysis of canopy asymmetry.

Along 10 transects, we randomly selected 555 *A. drepanolobium* trees 1–2 m tall. The lateral canopy spread of each focal tree was measured outwards from its rooting point along the four cardinal directions. In the cardinal quadrants surrounding each focal tree, we measured distance to the nearest neighbouring tree within 3 m, and noted which ant species occupied that neighbour. For *C. nigriceps* and *T. penzigi*, if the neighbouring ant was the same species as that on the focal tree, we experimentally transferred ants between trees and monitored their interactions to determine whether they belonged to different colonies. This assay was not conducted for *C. mimosae* and *C. sjostedti*, as these two species form extensive colonies on many trees, and aggressive interactions between adjacent trees are relatively rare. For each ant species, we conducted ANOVA with cardinal direction and the presence or absence of a neighbouring ant colony as predicting factors. Deviation from canopy symmetry, measured as the difference between lateral spread in one direction and the mean lateral extension for that tree, was the outcome variable.

Received 2 March; accepted 26 July 1999.

- Davidson, D. W. & McKey, D. The evolutionary ecology of symbiotic ant–plant relationships. *J. Hymenopt. Res.* **2**, 13–83 (1993).
- McKey, D. & Davidson, D. W. in *Biological Relationships between Africa and South America*; 37th Annual Systematics Symposium, St. Louis, Missouri, USA, October 4–6, 1990 (ed. Goldblatt, P.) 568–606 (Yale Univ. Press, New Haven, New Haven, 1993).
- Bentley, B. L. Extrafloral nectaries and protection by pugnacious bodyguards. *Ann. Rev. Ecol. Syst.* **8**, 407–427 (1977).
- Janzen, D. H. Coevolution of mutualism between ants and acacias in Central America. *Evolution* **20**, 249–275 (1966).
- Fiala, B. & Maschwitz, U. Food bodies and their significance for obligate ant-association in the tree genus *Macaranga* (Euphorbiaceae). *Botan. J. Linnean Soc.* **110**, 61–75 (1992).
- Risch, S. J. & Rickson, F. R. Mutualism in which ants must be present before plants produce food bodies. *Nature* **291**, 149–150 (1981).
- Fonseca, C. R. Herbivory and the long-lived leaves of an Amazonian ant-tree. *J. Ecol.* **82**, 833–842 (1994).
- Koptur, S. Experimental evidence for defense of *Inga* (Mimosoideae) saplings by ants. *Ecology* **65**, 1787–1793 (1994).
- Vasconcelos, H. L. Mutualism between *Maieta guianensis* Aubl., a myrmecophytic melastome, and one of its ant inhabitants: Ant protection against insect herbivores. *Oecologia* **87**, 295–298 (1991).
- Madden, D. & Young, T. P. Symbiotic ants as an alternative defense against giraffe herbivory in spinescent *Acacia drepanolobium*. *Oecologia* **91**, 235–238 (1992).
- Stapley, L. The interaction of thorns and symbiotic ants as an effective defence mechanism of swollen-thorn acacias. *Oecologia* **115**, 401–405 (1998).
- Janzen, D. H. Allelopathy by myrmecophytes: the ant *Azteca* as an allelopathic agent of *Cecropia*. *Ecology* **50**, 147–153 (1969).
- Davidson, D. W. & McKey, D. Ant plant symbioses: stalking the Chuyachaqui. *Trends Ecol. Evol.* **8**, 326–332 (1993).
- Davidson, D. W., Snelling, R. R. & Longino, J. T. Competition among ants for myrmecophytes and the significance of plant trichomes. *Biotropica* **21**, 64–73 (1989).
- Rico-Gray, V. & Thien, L. B. Effect of different ant species on reproductive fitness of *Schomburgkia tibicinis* (Orchidaceae). *Oecologia* **81**, 487–489 (1989).
- Vasconcelos, H. L. Ant colonization of *Maieta guianensis* seedlings, an Amazon ant-plant. *Oecologia* **95**, 439–443 (1993).
- Yu, D. W. & Davidson, D. W. Experimental studies of species-specificity in *Cecropia*–ant relationships. *Ecol. Monogr.* **67**, 273–294 (1997).
- Janzen, D. H. *Pseudomyrmex nigripilosa*: a parasite of a mutualism. *Science* **188**, 936–937 (1975).
- Davidson, D. W., Longino, J. T. & Snelling, R. R. Pruning of host plant neighbors by ants: An experimental approach. *Ecology* **69**, 801–808 (1988).
- Pellmyr, O., Leebens-Mack, J. & Huth, C. J. Non-mutualistic yucca moths and their evolutionary consequences. *Nature* **380**, 155–156 (1996).
- Bull, J. J. & Rice, W. R. Distinguishing mechanisms for the evolution of cooperation. *J. Theor. Biol.* **149**, 63–74 (1991).
- Fiala, B., Grunsky, H., Maschwitz, U. & Linsenmair, K. E. Diversity of ant–plant interactions: Protective efficacy in *Macaranga* species with different degrees of ant association. *Oecologia* **97**, 186–192 (1994).
- Yu, D. W. & Pierce, N. E. A castration parasite of an ant-plant mutualism. *Proc. R. Soc. Lond. B* **265**, 275–282 (1998).
- Coe, M. J. & Beentje, H. *A Field Guide to the Acacias of Kenya* (Oxford Univ. Press, Oxford, 1991).
- Young, T. P., Stubblefield, C. H. & Isbell, L. A. Ants on swollen-thorn acacias: Species coexistence in a simple system. *Oecologia* **109**, 98–107 (1997).
- Hocking, B. Insect associations with the swollen thorn acacias. *Trans. R. Entomol. Soc. Lond.* **122**, 211–255 (1970).
- Wood, W. F. & Chong, B. Alarm pheromones of the east African acacia symbionts: *Crematogaster mimosae* and *C. nigriceps*. *J. Georgia Entomol. Soc.* **10**, 332–334 (1975).
- Janzen, D. H. Interaction of the Bull's-Horn *Acacia* (*Acacia cornigera* L.) with an ant inhabitant (*Pseudomyrmex ferruginea* F. Smith) in Eastern Mexico. *Univ. Kansas Sci. Bull.* **6**, 315–558 (1967).
- Freund, R. J., Littell, R. C. & Spector, P. C. *SAS System for Linear Models* (SAS Institute, Cary, North Carolina, 1986).

Acknowledgements

We thank J. Lemboi, L. King, F. Lorugoi and F. Frei for field assistance, and the staff of Mpala Research Centre and Segera Ranch for logistical support. Mpala Ranch and Segera Ranch granted access to field sites. This research was supported by the National Science Foundation and by the Bridge Grant program at the University of California at Davis.

Correspondence and requests for materials should be addressed to M.S. (e-mail: mlstanton@ucdavis.edu).

Diet-dependent female choice for males with 'good genes' in a soil predatory mite

Izabela Lesna & Maurice W. Sabelis

Section for Population Biology, University of Amsterdam, Kruislaan 320, 1098 SM Amsterdam, The Netherlands

Female choice for mates with 'good genes' presupposes that some males have better genes than others¹. However, the resulting selection against inferior males causes such genetic variability to disappear. This paradox may be resolved when substantial variability is maintained at a balance between selection and mutation². Alternatively, populations may exhibit genetic polymorphisms maintained by frequency-dependent selection or hybrid vigour^{3–9}. Here we show that a local population of soil predatory mites exhibits genetic variation in preference for two prey species. We find that hybrids between selected preference lines are superior or inferior in population growth rate, depending on the composition of the diet. Finally, we show that females in this population mate disassortatively when hybrids are superior, but switch to assortative mating otherwise. Thus, mate choice varies with diet and is tuned to incorporate 'good genes' in the offspring, that is, genes that promote the population growth rate of the offspring on the same diet as that experienced by the parents. In this way, hybrid success and mate choice act together in maintaining or eliminating genetic polymorphism in local populations.

Polyphagy is thought to be widespread among predatory arthropods¹⁰, but the evidence comes from species- or population-wide pooling of data on prey consumption. To investigate genetic variation in food selection, we studied the small (<1 mm), soil-inhabiting predatory mite *Hypoaspis aculeifer* (Canestrini) (Acari: Laelapidae), which preys on fungivorous and herbivorous mites, insect eggs and larvae (notably beetles, flies, thrips and springtails), enchytraeid worms and nematodes^{11,12}. It reproduces by arrhenotokous parthenogenesis, which makes genetic analysis easy because isofemale lines can be created through mother–son mating^{11–13}.

We found that a local population of this predatory mite harbours genetic variation in prey choice with respect to two species of astigmatic mites: the bulb mite *Rhizoglyphus robini* Claparède (prey R), and the copra mite *Tyrophagus putrescentiae* (Schrank) (prey T). Predator and prey were collected from a very small area (0.25 m²) in a lily field. Both prey species were reared on yeast flakes at 15 °C, and predators were reared on prey T at 22 °C with 70% relative humidity. Prey preference was assessed by subjecting individual females to three two-choice tests. Those exclusively attacking prey R or T were used to start isofemale lines, yielding lines with opposite preferences after four generations¹² (Fig. 1). Analysis of preference in F₁ hybrids and F₁ × parent backcrosses showed that preference is inherited as if determined by a single gene (or several tightly linked genes)¹² (Figs

Table 1 Parameters of the predicted binomial distributions and G-test for goodness of fit to data

	Estimate of parameters p	G-statistic	Critical level, P (d.f.)
R -line	$p_{R-line} = 0.735$	4.437	$P \sim 0.11$ (2), n.s.
T -line	$p_{T-line} = 0.346$	21.222	$P \ll 0.001$ (2)
RT	$p_{hybrid} = (p_{R-line} + p_{T-line})/2 = 0.540$	6.520	$P \sim 0.090$ (3), n.s.
TR	$p_{hybrid} = (p_{R-line} + p_{T-line})/2 = 0.540$	9.398	$P \sim 0.025$ (3)
$(RT) \times R$	p_{hybrid} and p_{R-line}	0.366	$0.900 < P < 0.975$ (3), n.s.
$(RT) \times T$	p_{hybrid} and p_{T-line}	4.146	$0.100 < P < 0.500$ (3), n.s.
$(TR) \times R$	p_{hybrid} and p_{R-line}	2.947	$0.100 < P < 0.500$ (3), n.s.
$(TR) \times T$	p_{hybrid} and p_{T-line}	3.056	$0.100 < P < 0.500$ (3), n.s.

1 and 2, Table 1). The selected lines readily hybridize without loss of viability¹². Given this simple mode of inheritance and the small spatial scale at which the polymorphism is manifested, why did the natural population not diverge into various ecotypes with pre- and/or postmating barriers^{14,15}?

An important condition for divergence is that high population growth rates on one prey are achieved at the expense of that on another prey^{14,15}. This trade-off was examined by a two-way analysis of variance (ANOVA) on the relative growth rates of the predator populations ('base strain', selected lines and their hybrids) exposed to three different prey diets (**R**, **T** or a mixture) (Tables 2, 3). The line that prefers prey **R** (the R -line) was superior to that preferring prey

Table 2 Relative growth rates (mean $\pm$ s.e.m.) of different predator populations on three prey diets

Origin of predators	Relative growth rates (individuals per day) of predators fed on		
	prey R alone	prey T alone	prey R and T
Base strain	0.111 ± 0.006	0.114 ± 0.009	0.102 ± 0.011
R -line	0.132 ± 0.005	0.118 ± 0.008	-0.020 ± 0.012
T -line	0.086 ± 0.008	0.055 ± 0.018	0.045 ± 0.037
RT hybrids	0.091 ± 0.008	0.156 ± 0.003	0.039 ± 0.005
TR hybrids	0.122 ± 0.007	0.107 ± 0.005	0.065 ± 0.003

Values are based on four replicates of each treatment.

Table 3 Analysis of variance on growth rates of five predator populations on three prey diets

Source	d.f.	SS	MS	F-ratio	P
Population	4	0.019	0.005	16.685	1.9×10^{-8}
Diet	2	0.053	0.027	90.804	1.0×10^{-15}
Population $\times$ diet	8	0.040	0.005	17.297	2.0×10^{-11}
Error	45	0.013	0.0003		

SS, sum of squares; MS, mean of squares.

T (the T -line) in population growth rate, both on prey **R** and **T** ($P = 0.010$ and $P = 0.0010$, respectively, where P is the exact critical level). Moreover, the T -line and the R -line did not show significantly different population growth rates on prey **R** and **T**. Thus, high population growth rate on prey **R** does not imply lower population growth rate on prey **T** (and vice versa), and preference for **T** does not imply higher population growth rate on this prey. Food choice and population growth rate therefore correlate in one preference line (the R -line), but not in the other (the T -line). With no trade-off in performance between the two preference lines, ecotype differentiation seems unlikely, and, given the R -line superiority, one would expect genetic diversity to be eroded by natural selection and the R -preferring type to become fixed.

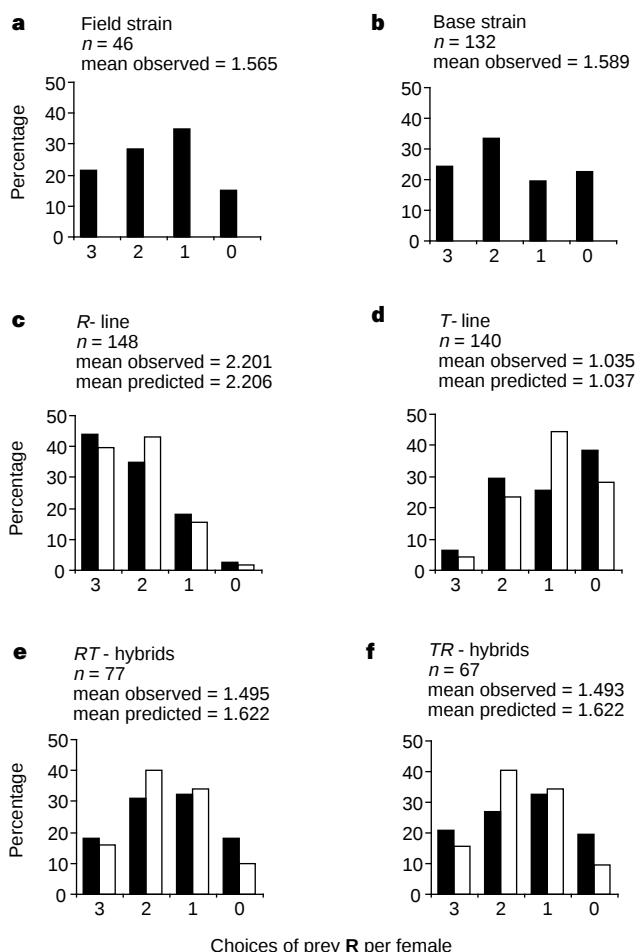


Figure 1 Observed (black columns) percentage of female predatory mites that, in three two-choice tests, attacked prey **R** 3, 2, 1 and 0 times. **a**, F_1 females reared from field-collected specimens; **b**, base strain; **c**, R -line; **d**, T -line; **e**, RT hybrids; **f**, TR hybrids. Predictions from binomial distributions (white columns) were not rejected, except for TR hybrids and T line (see Methods and Table 1); the latter may result from conditioning on prey **T**, offered between choice tests.

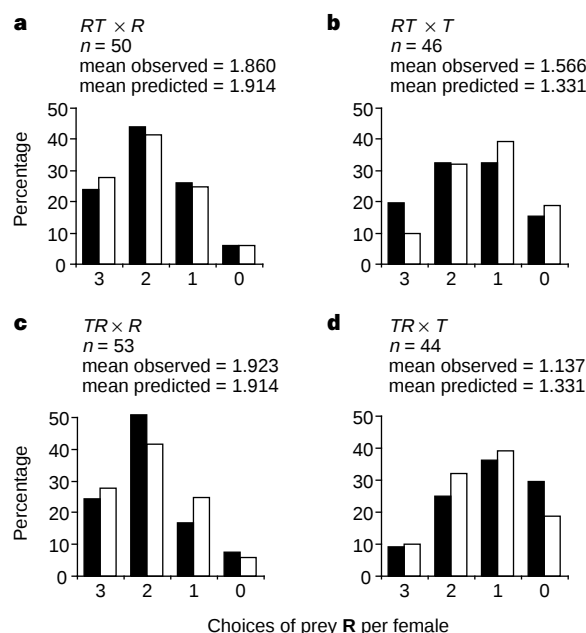


Figure 2 Histograms of attacks on prey **R** for $F_1 \times$ parent backcrosses. **a**, $RT \times R$; **b**, $RT \times T$; **c**, $TR \times R$; **d**, $TR \times T$. Crosses are labelled assuming monogenic inheritance of the preference trait in an arrhenotokous mite (haploid males, R , T ; diploid females, RR , TR , RT , TT ; first character indicates the mother's line and second character the father's line). Predictions from binomial distributions (white columns) were not rejected.

Table 4 Mate choice of female predators for R or T males on three prey diets.

Origin of females	Percentage of within-line matings of female predators fed on		
	prey R alone	prey T alone	prey R and T
<i>R</i> -line	70.6 ± 4.0 60.0–100	70.1 ± 2.5 57.1–83.3	26.9 ± 1.3 20.0–33.3
<i>T</i> -line	36.9 ± 2.6 25.0–55.0	33.3 ± 3.3 22.2–57.1	25.0 ± 3.5 0.0–40.0

Percentages (mean ± s.e.m. and range) of females that first mate with a male from their own line. Means and standard errors are based on 10 replicates of each treatment.

Table 5 ANOVA on ranked mate-choice data of two predator populations (lines) on three prey diets

Source	d.f.	SS	MS	<i>F</i> -ratio	<i>P</i>
Population	1	4,050.817	4,050.817	13.381	<i>P</i> < 0.001
Diet	2	6,687.475	3,343.738	22.091	<i>P</i> < 0.001
Population × diet	2	2,334.908	1,167.454	7.713	<i>P</i> < 0.025
Error	54	4,787.3	88.654		
Total	59	17,860.5	302.72		

Analysis was by Scheirer–Ray–Hare extension of the Kruskal–Wallis test. If the dependent variable had been defined as percentage of matings with males of the *R*-line (instead of own line), the effect of population on variation in mate choice would not have been significant.

To investigate whether coexistence is facilitated when the two prey species occur together, we measured population growth of the lines on mixed prey. This showed that the *R*-line performed much worse on the prey mixture (with negative population growth rates in all four replicates) than on each of the two prey alone (*P* = 0.00015 for prey **R**, *P* = 0.00015 for prey **T**) (Table 2). The bad performance of the *R*-line on the prey mixture probably results from digestive problems (I.L., unpublished data). However, the *T*-line on the prey mixture, although more variable, did not perform significantly better or worse than on either prey **R** (*P* = 0.080, not significant (n.s.)) or **T** alone (*P* = 0.995, n.s.). In conclusion, the *T*-line outperformed the *R*-line on the mixture of prey (*P* = 0.00083), whereas the reverse was true on each of the two prey species offered alone. Hence, microspatial and temporal variability in prey composition may favour the coexistence of preference types^{16,17}. Indeed, because the base strain showed higher population growth rates on the prey mixture than did the *R*-line (*P* = 0.00015) or the *T*-line (*P* = 0.0006) (Table 2), it may contain additional phenotypic variability owing to hybrids or other genotypes.

We investigated the first of the possibilities by assessing population growth rates of hybrids on each prey species alone and on the prey mixture. When fed **R**, the *RT* and *TR* hybrids together performed intermediately between the selected lines, with the *RT* hybrids closer to the *T*-line and the *TR* hybrids closer to the *R*-line (Table 2). When fed **T**, the (*RT* plus *TR*) hybrids showed similar population growth rates to the *R*-line (*P* = 0.069, n.s. for *RT* hybrids; *P* ~ 1, n.s. for *TR* hybrids) and higher than the *T*-line (*P* = 0.00015 for *RT* hybrids; *P* = 0.006 for *TR* hybrids) (Table 2). When offered a mixture of **R** and **T**, both hybrids performed as well as the *T*-line (*P* ~ 1 for *RT* hybrids, n.s.; *P* = 0.759 for *TR* hybrids, n.s.), but outperformed the *R*-line (*P* = 0.001 for *RT*; *P* = 0.00015 for *TR*) (Table 2). Why producing fit offspring on the prey mixture should be achieved by hybridization, and so be subject to genetic constraints, is unclear, although a mixed prey diet is hard to digest for the *R*-line but not for the hybrids or the *T*-line (I.L., unpublished data). A systematic search for ‘masters of all trades’ may provide more insight into the genetic constraints, but this is not our aim here.

Thus, diet determines whether hybrids have higher, equal or lower population growth rates than one or both selected lines. This indicates that hybrid vigour may be another mechanism that maintains genetic diversity within local populations of predatory mites where prey **R** and **T** occur together, but also in an environment with **T** alone, because hybrids perform at least as well as the best of the two parental lines (the *R*-line). However, when **R** is the only prey

available, the population is expected to become fixed with respect to preference for prey **R** (provided that no other genotypes do even better). This expectation was borne out in a long-term culture of the base strain on prey **R**, where **R**-preferring types gradually became dominant¹².

If hybrid vigour is to provide an additional explanation for the maintenance of local genetic diversity, then mating structure in local populations must be such that mating occurs between genotypes with opposite prey preferences. This was investigated in laboratory experiments on mate choice. Young virgin females were offered a choice between *R*-line and *T*-line males. To speed up mating and to offer individual females a choice between potential partners, each replicate experiment was carried out with groups of eight females and eight males from either line. The experiments were performed in the presence of prey (**R**, **T** or a mixture) on which females were reared for two days before the trial. Males were always reared on prey **T**. On all three prey diets, *T*-line females mated disassortatively, that is, with males of the *R*-line, whereas *R*-line females mated assortatively with males of their own line when fed on **R** or **T** alone. When fed on the prey mixture, however, they did not mate assortatively, but mated preferentially with males of the *T*-line (negative assortative mating). Thus, under dietary conditions in which hybrids are clearly superior, mating occurs predominantly between the two contrasting preference types, and when parental lines are superior, assortative mating predominates. The mixture of prey **R** and **T** therefore promotes polymorphisms by a combination of hybrid superiority and a negative correlation between mate choice and food choice. On a diet of **T**, the same applies because the *T*-line mates disassortatively and the hybrids do better than the *T*-line and at least as well as the *R*-line. On a diet of **R**, polymorphisms are expected to vanish because of the combination of superior population growth rate and assortative mating in the *R*-line. As referred to earlier, this predication was borne out in a long-term culture of the base strain on prey **R**¹². This shows how diet-dependent mate choice determines the maintenance or loss of genetic variability within local populations. In retrospect, these results explain why the field population at the site of collection (a roughly 20:1 mixture of prey **R** and **T**) was polymorphic and why the base strain retained its polymorphism when reared in the laboratory on prey **T**. Under natural conditions, prey composition changes over a timescale of days to weeks, whereas the time for mate finding and mating takes minutes to hours. A female choosing a mate can therefore estimate the conditions to be experienced by her offspring from the conditions at the time of mating.

Mating systems that promote hybridization among offspring by mating between contrasting parental lines have been considered unlikely to be a general mechanism⁹. However, our result—that female predatory mites mate with males from another preference line when producing hybrids is advantageous—provides evidence in favour of this, as do some studies on flies^{18,19}, sparrows²⁰ and mice^{21–26} (but see ref. 27 for a critical evaluation). This is because males of the predatory mites we studied are haploid, and so cannot have been selected by females for their heterozygosity⁹. Female predatory mites invest heavily in eggs, so they are the choosy sex and must have selected males for their effect on the offspring, not because they were particularly vigorous. Moreover, as shown for the *R*-line, females switch to mating with males of their own line when hybrid offspring is inferior. Because male predators were all reared on the same diet (**T**) before the mate-choice experiment, this switch cannot be mediated by an effect of diet on male quality. Taken together, our data indicate that females select males with good genes for their offspring, rather than for male vigour, but a definitive conclusion awaits similar studies with more than two lines. □

Methods

Genetics of prey choice

In September 1991, 150 predatory mites were collected from 0.25 m² sandy soil containing

20 lily bulbs in a field near Breezand in the Netherlands. These mites founded the lab-reared base strain. For selection experiments, starting in January 1995, 132 female deutonymphs (near moulting) were placed individually in vials with five female prey T. After one day, matured predators were starved for three days with access to water (no mortality). After putting one female predator, one female and one protonymph of each prey species in each vial, they were observed twice per minute until an attack on prey R or T. This procedure (feeding for 1 day, starvation for 3 days; choice test) was repeated three times per individual. Predator females that selected three times prey R or T were mated with their own sons to create isofemale lines, which were fed on T. This selection procedure was repeated for four generations, yielding one line preferring R (the R-line) and one preferring T (the T-line). As above, R-line (148) and T-line (140) females were subjected to three prey-choice tests. For each line, ten females were cross-bred with males of the other line. Hybrid females with R-line mothers (77) and those with T-line mothers (67) were subjected to three prey-choice tests, as above. Ten F_1 hybrid females were then backcrossed with males from each parental line. From offspring of each of the four types of $F_1 \times$ parent backcrosses, ~50 young females were subjected individually to three prey-choice tests. Frequency distributions of prey choices are shown in Fig. 1 (base strain, R-line, T-line, F_1 -hybrids) and Fig. 2 ($F_1 \times$ parent backcrosses). Figure 1 also shows prey-choice frequencies assessed in progeny of females collected in September 1998 (from near the location sampled in 1991).

Estimates of p , the probability of choosing prey R, were obtained from observed choice distributions of the parental lines according to $p = (3p_{3R} + 2p_{2R} + p_{1R})/3$. Assuming monogenic inheritance without dominance, $p_{R\text{-line}}$ and $p_{T\text{-line}}$ were averaged to predict $p_{RT} = p_{TR}$. The estimates of $p_{R\text{-line}}$, $p_{T\text{-line}}$ and $p_{RT} = p_{TR}$ were used to predict frequency distributions of choices for the four $F_1 \times$ parent backcrosses (for a haplodiploid, a 1:1 ratio of homozygote and heterozygote daughters is expected). Observed frequency distributions of prey choices were tested for goodness of fit to predicted binomial distributions, using a G-test with Williams's correction²⁸ (d.f. = 4 - 1 - 1 = 2, when p is estimated from the observed distribution under test; d.f. = 4 - 1 = 3, when p is estimated from an independent distribution) (Table 1).

Population growth on excess food

Predator population size (N) was recorded three times each week for 4 weeks under ample prey supply (predator:prey ratio maintained at 1:100) (22 °C, 70% relative humidity). Predators were supplied with either prey R alone, T alone, or a 1:1 mixture of R and T, in vials (7 cm diameter, 7 cm high; lid with pinholes covered by mite-proof gauze). All populations were started with a representative sample of 25 individuals including all mobile stages, taken from growing cultures (3–4 weeks old) of either base strain, R-line, T-line or (RT or TR) hybrids. Because $\ln(N)$ increased or decreased linearly over the observation period (t), relative population growth rates, $r = \ln(N_t/25)/t$ were calculated per replicate. Each treatment (combination of predator origin and diet) was replicated 4 times (Table 2). Two-way ANOVA was applied with r as the dependent variable and with predator populations (base strain, R-line, T-line and RT-, TR-hybrids) and prey diets (R, T, mixture) as factors (Table 3). Post-hoc comparisons were made using the T-method ($\alpha = 0.05$)²⁸.

Mate choice

Female deutonymphs from cultures of R-line or T-line predators were put in groups of 25 per vial. After sexual maturation (1–2 days) on prey R alone, T alone or a 1:1 mixture of R and T, eight females were put in each vial (3 cm diameter, 4 cm high) with eight young R-line males, eight young T-line males and the same prey as used for conditioning the females. The males, reared invariably on prey T, were marked dorsally using water paint. Virgin females were given 0.5 h for mate selection. Vials were inspected once per minute (mating time >5 min). Mate choice per vial was calculated as the percentage of females selecting an own-line male. Each treatment (female origin and diet) was replicated 10 times (Table 4). Two-way ANOVA was applied on ranked data, using the Scheirer–Ray–Hare extension of the Kruskal–Wallis test²⁸ with the percentage of within-line matings as the dependent variable and predator populations (R-line, T-line) and diets (R, T, mixture) as factors (Table 5).

Received 1 June; accepted 26 July 1999.

- Andersson, M. *Sexual Selection* (Princeton Univ. Press, NJ, 1994).
- Bulmer, M. G. Maintenance of genetic variability by mutation–selection balance. *Genome* **31**, 761–767 (1989).
- Falconer, D. S. & McKay, R. *Introduction to Quantitative Genetics* 3rd edn (Longman, New York, 1996).
- Felsenstein, J. Skepticism towards Santa Rosalia, or why there are so few kinds of animals. *Evolution* **35**, 124–138 (1981).
- Hoekstra, R. F., Bijlsma, R. & Dolman, A. J. Polymorphism from environmental heterogeneity: Models are only robust if the heterozygote is close in fitness to the favoured homozygote in each environment. *Genet. Res.* **45**, 299–314 (1985).
- García-Dorado, A. The effect of niche preference on polymorphism protection in a heterogeneous environment. *Evolution* **40**, 936–945 (1986).
- Meeus, T. de, Michalakos, Y., Renaud, F. & Olivieri, I. Polymorphism in heterogeneous environments, evolution of habitat selection and sympatric speciation: soft and hard selection models. *Evol. Ecol.* **7**, 175–198 (1993).
- Jones, J. S. Can genes choose habitats? *Nature* **286**, 757–758 (1980).
- Brown, J. L. A theory of mate choice based on heterozygosity. *Behav. Ecol.* **8**, 60–65 (1997).
- Sabelis, M. W. in *Natural Enemies, The Population Biology of Predators, Parasites and Diseases* (ed. Crawley, M. J.) 225–264 (Blackwell Scientific, Oxford, 1992).
- Lesna, I., Sabelis, M. W. & Conijn, C. G. M. Biological control of the bulk mite, *Rhizoglyphus robini*, by the predatory mite, *Hypoaspis aculeifer*, on lilies: Predator–prey interactions at various spatial scales. *J. Appl. Ecol.* **33**, 369–376 (1996).

- Lesna, I. Bulb mite biocontrol: evolutionary genetics of prey choice in soil predators. Ph.D. Thesis, Univ. Amsterdam (1998).
- Witalinski, W. Sperm competition in the Acari. *Exp. Appl. Acarol.* **23**, 455–465 (1999).
- Maynard Smith, J. Sympatric speciation. *Am. Nat.* **100**, 637–650 (1966).
- Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19**, 207–233 (1988).
- Chesson, P. L. in *Community Ecology* (eds Diamond, J. & Case, T. J.) 240–254 (Harper and Row, New York, 1986).
- Chesson, P. A need for niches? *Trends Ecol. Evol.* **6**, 26–28 (1991).
- Serradilla, J. M. & Ayala, F. J. Alloprocytic selection: A mode of natural selection promoting polymorphism. *Proc. Natl Acad. Sci. USA* **80**, 2022–2025 (1983).
- Gilburn, A. A. & Day, T. H. Evolution of female choice in seaweed flies: Fisherian and good genes mechanisms operate in different populations. *Proc. R. Soc. Lond. B* **255**, 159–165 (1994).
- Houtman, A. M. & Falls, J. B. Negative assortative mating in the white-throated sparrow: the effect of mate choice and intrasexual competition. *Anim. Behav.* **48**, 377–383 (1994).
- Lenington, S. The t-complex: a story of genes, behavior, and populations. *Adv. Study Behav.* **20**, 51–86 (1991).
- Coopersmith, C. B. & Lenington, S. Preferences of female mice for males whose t-haplotype differs from their own. *Anim. Behav.* **40**, 1179–1181 (1990).
- Lenington, S., Coopersmith, C. B. & Erhart, M. Female preference and variability among t-haplotypes in wild house mice. *Am. Nat.* **144**, 766–784 (1994).
- Potts, W. K., Manning, C. J. & Wakeland, E. K. The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. *Phil. Trans. R. Soc. Lond. B* **346**, 369–378 (1994).
- Penn, D. J. & Potts, W. K. The evolution of mating preferences and major histocompatibility complex genes. *Am. Nat.* **153**, 145–164 (1999).
- Penn, D. J. & Potts, W. K. MHC-disassortative mating preferences reversed by cross-fostering. *Proc. R. Soc. Lond. B* **265**, 1299–1306 (1998).
- Edwards, S. V. & Hedrick, P. W. Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol. Evol.* **13**, 305–311 (1998).
- Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, New York, 1995).

Acknowledgements

We thank M. Boerlijst, H. Breeuwer, S. Elliot, A. Janssen, R. Lingeman, Y. Michalakos, N. van Straalen, F. Vala and A. Weeks for discussion and comments on the manuscript. This work was funded by the Technology Foundation (S.T.W.) and the Netherlands Organization of Scientific Research (NWO).

Correspondence and requests for materials should be addressed to I.L. (e-mail: lesna@bio.uva.nl).

fMRI evidence for objects as the units of attentional selection

Kathleen M. O'Craven*†‡, Paul E. Downing*‡ & Nancy Kanwisher*‡

* Massachusetts General Hospital NMR Center, Building 149 13th St., Charlestown, Massachusetts 02129, USA

† Bunting Institute at Radcliffe College, 34 Concord Ave, Cambridge, Massachusetts 02138, USA

‡ MIT Department of Brain and Cognitive Sciences, 77 Massachusetts Ave., Cambridge, Massachusetts 02139, USA

Contrasting theories of visual attention emphasize selection by spatial location¹, visual features (such as motion or colour)^{2–4} or whole objects^{5,6}. Here we used functional magnetic resonance imaging (fMRI) to test key predictions of the object-based theory, which proposes that pre-attentive mechanisms segment the visual array into discrete objects, groups, or surfaces, which serve as targets for visual attention^{5–9}. Subjects viewed stimuli consisting of a face transparently superimposed on a house, with one moving and the other stationary. In different conditions, subjects attended to the face, the house or the motion. The magnetic resonance signal from each subject's fusiform face area¹⁰, parahippocampal place area¹¹ and area MT/MST¹² provided a measure of the processing of faces, houses and visual motion, respectively. Although all three attributes occupied the same location, attending to one attribute of an object (such as the motion of a moving face) enhanced the neural representation not only of that attribute

but also of the other attribute of the same object (for example, the face), compared with attributes of the other object (for example, the house). These results cannot be explained by models in which attention selects locations or features, and provide physiological evidence that whole objects are selected even when only one visual attribute is relevant.

A central question in current theories of visual attention is whether the units of attention are locations, features, objects, or a combination of these. All of these theories agree that attention entails selection of some associated but irrelevant information, but they make different predictions about which irrelevant information will be automatically selected along with the target. fMRI has advantage over behavioural tests of these predictions^{13,14} because it allows the strength of specific neural representations to be measured directly and simultaneously without disrupting the selection task under investigation. Specifically, the magnetic resonance signal from the fusiform face area (FFA), which responds more strongly to faces than to other objects¹⁰, provides a measure of the neural processing of a face stimulus; the signal from the parahippocampal place area (PPA), which responds more strongly to places and houses than to other objects¹¹, provides a measure of the neural processing of a house stimulus; and the signal from area MT/MST¹² provides a measure of the neural processing of motion.

We asked two questions for which the different models of attention make contrasting predictions. First, is processing enhanced for all visual attributes at the attended location, as predicted only by the space-based theory? Second, does attention to one attribute of an object automatically entail processing of task-irrelevant visual attributes of the same object, as predicted by object-based models?

Each subject was scanned on a General Electric 3T Signa scanner on two runs of a functional localizer that enabled precise and independent identification of the voxels comprising his or her fusiform face area, parahippocampal place area and area MT/MST, as described previously^{10–12}. These three sets of voxels served as the regions of interest (ROIs) to be used in the analysis of the data from the main experiments.

In our first experiment, seven subjects were run on 6–8 scans, each consisting of one block of each of the six conditions created by crossing two stimulus types with three attentional conditions. In half the blocks, subjects viewed a series of stationary houses transparently superimposed on faces that oscillated along one of four axes, with a new display presented every 1.4 s (Fig. 1). In the other half of the blocks, the stimulus was identical except that the house was moving and the face was stationary. The subjects' attention was directed to a different stimulus attribute in each block by instructing them to monitor the series of displays for



Figure 1 Sample stimuli. A sample stimulus from the first experiment is shown on the left. For the second experiment the faces were enlarged to more closely match the region of space occupied by the house in each pair. An example is shown at the right, with the face displaced (in this case, to the right of fixation) as required for the position-repetition detection task. In each stimulus, either the face or the base moved back and forth along one axis.

consecutive repetitions of either the face, the house or the direction of motion.

For each of the three ROIs (FFA, PPA and MT/MST) in each subject, we computed the average per cent increase in magnetic resonance signal from the fixation baseline for each of the six experimental conditions. (For raw time course of per cent signal change over the period of the scan for each of the three ROIs, see Supplementary Information.) The averages of these values across subjects are given in Fig. 2. The resulting data set allowed us to confirm two fundamental predictions of object-based theories of attention. First, in each region the change in magnetic resonance signal was greater when subjects attended to the preferred attribute for that cortical region (for example, faces for the FFA) than when they attended to a different attribute of the same display (see comparisons shown in the grey boxes in Fig. 2). This effect was reliable in each of three ROIs tested: FFA ($F(1,6) = 31.9$, $P < 0.001$), PPA ($F(1,6) = 50.7$, $P < 0.0001$), and MT/MST ($F(1,6) = 19.5$, $P < 0.005$). This attentional modulation cannot be accounted for by a purely spatial model of attentional selection, because all of the relevant information occupied the same location.

Second, we tested the central claim of object-based theories: task-irrelevant attributes of an attended object will be selected along with the task-relevant attribute, even when these attributes are independent (as the form and motion attributes were here). This prediction was confirmed for each extrastriate area (see comparisons shown in the black rectangles in Fig. 2). When the subject attended to motion, a higher signal change was observed in the FFA when the faces moved than when the houses moved (with faces stationary) ($F(1,6) = 8.5$, $P < 0.05$). The complementary result was obtained in the PPA: the signal change was greater during the motion task when the house moved than when the face moved ($F(1,6) = 42.5$, $P < 0.001$), even though both faces and houses were present in each condition. Finally, in MT/MST, the signal change was greater when subjects attended to the face or house when they were moving than when they were stationary ($F(1,6) = 16.0$, $P < 0.01$). This effect did not depend on whether subjects were attending to faces or houses ($F(1,6) < 1$, n.s.).

These results cannot be explained by space- or feature-based models of attention alone. Instead they show that, even when the task requires only that subjects select a given visual attribute, both attributes of the attended object are automatically selected.

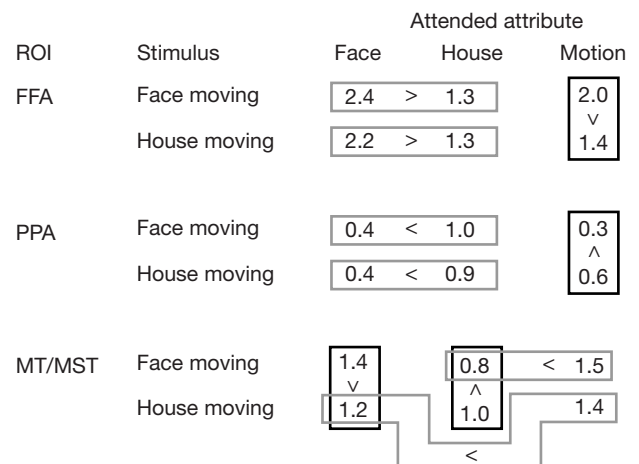


Figure 2 Design and results of Experiment 1. For each condition, the average per cent signal change across subjects is given for each of the three ROIs. The grey boxes show the greater response in each extrastriate ROI to the attended attribute compared with the unattended attribute, which cannot be accounted for by a purely location-based theory of attention. The black boxes show the greater response to the irrelevant attribute of the attended object compared with the unattended attribute of the unattended object, as predicted only by object-based theories of attention (see text for statistics).

Furthermore, it would be very difficult to account for these results in terms of smooth-pursuit eye movements; even if subjects ignored our instructions to fixate, any effort to track the moving object would have been ineffective because it reversed direction every 135 ms. Thus, our findings are best explained by object-based theories of attention. However, one remaining possibility is that the blocked design may have encouraged subjects to use the task-irrelevant attribute to aid in selection of the relevant attribute of the same object.

The second experiment precluded this possibility by using an event-related design¹⁵ in which we randomly intermixed trials containing the two stimulus types (moving faces with stationary houses, or moving houses with stationary faces). Four subjects performed a consecutive repetition-detection task in different scans on either the direction of motion of the moving object or the position of the stationary object (which was offset slightly in each stimulus to the left, right, top or bottom of fixation). This design prevented subjects from predicting from trial to trial whether the task-relevant object (the moving or stationary object) would be a face or a house. Thus, attention to the task-irrelevant dimensions could provide no benefit to performance of the required task.

Consistent with previous studies, the evoked response to the stimulus peaked 4–6 s after stimulus onset. We therefore took the average of the three corresponding time points as our measure of response magnitude (Fig. 3). In the motion task, the neural representation of the form (face or house) of the moving object was enhanced, confirming our finding that the processing of the irrelevant attribute of the attended object is increased ($F(1, 3) = 143.6$, $P < 0.001$). For the position task, we found the opposite pattern: the representation of the stationary face or house was enhanced relative to the moving stimulus ($F(1, 3) = 17.1$, $P < 0.05$). The significant interaction of task $\times$ stimulus $\times$ cortical area ($F(1, 3) = 40.8$, $P < 0.01$) eliminates an account of the data in terms of generally enhanced processing of either moving or stationary stimuli. To ensure that baseline effects were not responsible for any of these results, the same analyses were conducted on the first three time points of the evoked response. No significant effects were found in these analyses (all $P > 0.10$).

Thus, Experiment 2 provides strong support for the key prediction of object-based theories of attention. During a task that requires selection of only one visual attribute, the representation of the other attributes of the same object are automatically enhanced, even if their combination with the target dimension is unpredictable and uninformative. This result rules out any explanation of the data in terms of strategic use of the irrelevant dimension in attentional selection. Furthermore, enhancement of the irrele-

vant attribute occurred even though these attributes (faces and houses) were never task relevant for the subjects in this experiment.

Our data provide, to our knowledge, the first neural evidence that objects serve as the units of attention even when the selection of objects is not required by the task. However, it is unlikely that objects are the only units of attentional selection. The ability to select features and/or feature dimensions is supported by evidence from behavioural^{2,4}, imaging^{3,16,17}, single-unit¹⁸ and event-related potential¹⁹ studies. Our finding that the neural response to the task-relevant attribute was stronger than the response to the task-irrelevant attribute of the attended object (Fig. 2) may reflect such a feature-based effect, although it may also reflect the involvement of extrastriate areas in working memory²⁰ or perceptual decision-making²¹.

Spatial locations probably also serve as the units of selection under some circumstances²². Several recent fMRI studies have shown attentional modulations of the response to visual stimuli in retinotopic cortex, including area V1^{23–25}. However, none of these studies unconfounded object-based selection from location-based selection, because in all such studies each location tested was occupied by a distinct object. Our experiments solve this problem by superimposing two objects in the same location, to preclude location-based selection. Two other studies^{16,26} have also shown attentional modulation of objects or textures superimposed in the same location, consistent with object-based attention. However, our study adds a critical new condition enabling us to measure the response to the task-irrelevant attribute of the attended object.

The 'biased-competition' model of attention²⁷ suggests the following interpretation of our data. The instruction to attend to a particular dimension (such as motion) results in a top-down bias signal²⁸ which enhances responses in the extrastriate area coding that dimension (such as MT/MST). This increased response to the attended attribute causes an enhancement of the neural response to the other attributes of the same object (for example, the face, represented in the FFA). This enhancement occurs even when the task does not require subjects to determine which object is moving (that is to bind the visual attributes of form and motion).

The biased-competition model leaves two crucial questions unanswered. First, how are the correct visual attribute bindings determined in the first place? Feature-integration theory²⁹ offers a solution, but it relies on location information and so would not be effective for the present case of superimposed objects. Psychophysical and modelling results³⁰ indicate that, for transparent motion, the solution to the binding problem may require information present only at very early stages of the visual pathway; the same is likely to be true for binding of other independent visual dimensions such as colour and disparity. Second, how do these bindings (once computed) lead visual information represented in one extrastriate neural population (such as MT/MST) to affect the response of a neural population in a remote visual area (such as FFA)? One possibility is that recurrent feedback between each extrastriate area and early stages of the visual pathway is central to the competitive interactions that enable representations of objects to be constructed and to become the units of attention and perceptual awareness. □

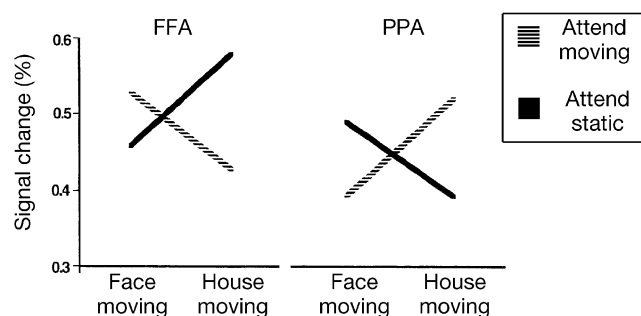


Figure 3 Means of the peak evoked responses from Experiment 2 in the FFA and PPA as a function of task (attend moving or attend static) and stimulus (face moving or house moving). Even though subjects never had to attend to faces or houses in the entire experiment, and could not predict which stimulus would move on a given trial, magnetic resonance responses were greater to the task-irrelevant attribute of the attended object than to the irrelevant attribute of the unattended object, as predicted by object-based theories of attention.

Methods

Subjects and task

Seven normal young adults served as subjects in Experiment 1, and four in Experiment 2. Data from two additional subjects from Experiment 1 and one from Experiment 2 were excluded because of failure to identify one of the three ROIs in the localizer scans. Each subject gave informed consent before participation.

Each stimulus was composed by transparently superimposing one of eight greyscale front-view photographs of young white male faces on one of eight greyscale photographs of houses. The stimulus subtended $\sim 10^\circ$ of visual angle.

In Experiment 1, each trial consisted of a 675-ms stimulus presentation, during which the moving item oscillated (2 cycles) along a straight path on one of four non-cardinal axes. The maximal displacement caused by the motion was less than 10% of the width of the image. The inter-trial interval was 725 ms. Each scan consisted of six 28-s task blocks

(attend face, attend house or attend motion $\times$ face moving, house moving) alternating with 10-s blocks in which only a fixation point was present. The initial, middle and final fixation blocks were extended to 30 s to provide a stable baseline measure. The block order was reversed on half of the runs. A cue word ('face', 'house' or 'motion') appearing immediately before each block indicated which task should be performed next. Subjects were instructed to fixate on a central dot that remained present throughout the experiment, and to press a button to indicate a consecutive repetition of the designated attribute. Only stimuli on the task-relevant dimension were repeated, with a probability of 0.125.

The stimuli and task in Experiment 2 were identical to Experiment 1 except as follows. Subjects monitored for consecutive repetitions of either motion direction or position relative to the fixation point in separate scans. Each stimulus presentation lasted 375 ms, with a fixation dot present at all times. On each trial, one item was stationary and was displaced $\sim 1.0^\circ$ in one of four directions from the centre. The other item moved along one of the four motion paths, making one excursion out and back to the centre. Each scan contained three trial types, each 2 s long: static house with moving face; static face with moving house; and fixation. Twenty-four trials from each condition were tested in each scan. The presentation sequence was uniquely randomized for each scan. Trials from each condition were preceded equally often by trials from each of the three conditions. Repetitions on each stimulus dimension occurred with probability 0.125, with no correlation across dimensions.

Imaging and data analysis

Scanning was done at the MGH NMR Center in Charlestown, Massachusetts, on a 3 Tesla General Electric Signa MR scanner modified by ANMR to perform echo-planar imaging. A gradient echo pulse sequence (TR = 2s for Experiment 1 and localizer scans; TR = 1s for Experiment 2; TE = 30 ms) was employed. Eight near-coronal slices (parallel to the brainstem, 7 mm thick with 3.125×3.125 mm in-plane resolution) were collected with a custom surface coil (built by T. Vaughan) to enhance the signal-to-noise ratio in the posterior brain regions under investigation.

The FFA was defined as the set of all contiguous voxels in the fusiform gyrus that responded more strongly to faces than to houses; the PPA was defined as all contiguous voxels in the parahippocampal region that responded more strongly to houses than to faces; and MT/MST was defined as all contiguous voxels in the occipito-temporo-parietal junction that responded more strongly to moving faces, houses and dots than to stationary ones. For all three analyses we used a threshold of $P < 10^{-4}$ (uncorrected) on a Kolmogorov-Smirnov test.

For Experiment 1, the mean per cent signal change was computed by subject for each condition in each ROI. To compensate for haemodynamic lag and for mixtures of effects at the boundaries of epochs, the first data point of each epoch was assigned to the previous epoch, and the next two data points were omitted from the analysis.

For Experiment 2, the fMRI signal from the FFA and PPA was extracted for each scan separately. Twelve time-points (12 s) of data were averaged by condition, beginning from the onset of each trial. The data were converted to per cent signal change relative to the corresponding timepoint on fixation trials, and mean per cent signal changes were then calculated for each of the four conditions of interest for each subject. The peak of the evoked response (the mean of timepoints 4, 5 and 6) was analysed with ROI (FFA or PPA), task (attend motion or attend position) and stimulus (house moving or face moving) as factors.

In all other respects, the methods used here were as reported^{10,11,16}.

Received 24 June; accepted 2 August 1999.

- Posner, M. I. Orienting of attention. *Q. J. Exp. Psychol.* **32**, 3–25 (1980).
- Treisman, A. M. Strategies and models of selective attention. *Psychol. Rev.* **76**, 282–299 (1969).
- Corbetta, M. *et al.* Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* **248**, 1556–1559 (1990).
- Wolfe, J. M., Cave, K. R. & Franzel, S. L. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perf.* **15**, 419–433 (1989).
- Duncan, J. Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* **113**, 501–517 (1984).
- He, Z. J. & Nakayama, K. Visual attention to surfaces in three-dimensional space. *Proc. Natl Acad. Sci. USA* **24**, 11155–11159 (1995).
- Vandenbergh, R. *et al.* Attention to one or two features in left or right visual field: a positron emission tomography study. *J. Neurosci.* **17**, 3739–3750 (1997).
- Roelfsema, P. R., Lamme, V. A. & Spekreijse, H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* **395**, 376–381 (1998).
- Driver, J., Baylis, G. C., Goodrich, S. J. & Rafal, R. D. Axis-based neglect of visual shapes. *Neuropsychologia* **32**, 1353–1365 (1994).
- Kanwisher, N., McDermott, J. & Chun, M. The fusiform face area: A module in human extrastriate cortex specialized for the perception of faces. *J. Neurosci.* **17**, 4302–4311 (1997).
- Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
- Tootell, R. B. *et al.* Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* **15**, 3215–3230 (1995).
- Eriksen, B. A. & Eriksen, C. W. Effects of noise letters upon the identification of target letters in a nonsearch task. *Percept. Psychophys.* **16**, 143–149 (1974).
- Driver, J. & Baylis, G. C. Movement and visual attention: The spotlight metaphor breaks down. *J. Exp. Psychol. Hum. Percept. Perf.* **15**, 448–456 (1989).
- Rosen, B. R., Buckner, R. L. & Dale, A. M. Event-related functional MRI: Past, present, and future. *Proc. Natl Acad. Sci. USA* **95**, 773–780 (1998).
- O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A. & Savoy, R. L. Voluntary attention modulates fMRI activity in human MT-MST *Neuron* **18**, 591–598 (1997).

- Beauchamp, M. S., Cox, R. W. & DeYoe, E. A. Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *J. Neurophys.* **78**, 516–520 (1997).
- Treue, S. & Trujillo, J. C. M. Feature-based attention influences motion processing in macaque visual cortex. *Nature* **399**, 575–579 (1999).
- Hillyard, S. A. & Munte, T. F. Selective attention to color and location: An analysis with event-related brain potentials. *Percept. Psychophys.* **36**, 185–198 (1984).
- Courtney, S. M., Ungerleider, L. G., Keil, K. & Haxby, J. V. Transient and sustained activity in a distributed neural system for human working memory. *Nature* **386**, 608–611 (1997).
- Salzman, C. D. & Newsome, W. T. Neural mechanisms for forming a perceptual decision. *Science* **264**, 231–237 (1994).
- Mangun, G. & Hillyard, S. A. Allocation of visual attention to spatial location: Tradeoff functions for event related brain potentials and detection performance. *Percept. Psychophys.* **47**, 532–550 (1990).
- Brefczynski, J. A. & DeYoe, E. A. A physiological correlate of the 'spotlight' of visual attention. *Nature Neurosci.* **4**, 370–374 (1999).
- Somers, D. C., Dale, A. M., Seiffert, A. E. & Tootell, R. B. Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl Acad. Sci. USA* **96**, 1663–1668 (1999).
- Gandhi, S. P., Heeger, D. J. & Boynton, G. M. Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl Acad. Sci. USA* **96**, 3314–3319 (1999).
- Valdes-Sosa, M., Bobes, M., Rodriguez, V. & Pinilla, T. Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *J. Cogn. Neurosci.* **10**, 137–151 (1998).
- Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
- Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R. & Ungerleider, L. G. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Neuron* **22**, 751–761 (1999).
- Treisman, A. M. & Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980).
- Qian, N., Andersen, R. A. & Adelson, E. H. Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *J. Neurosci.* **14**, 7357–7366 (1994).

Supplementary information is available at Nature's World-Wide Web site (<http://www.nature.com>.) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank B. Rosen and others at the MGH-NMR Center for support; H. Wyatt for consultation on eye movements; and B. Anderson, J. Driver, R. Epstein, J. Mausell, M. Potter, G. Rainer, F. Tong, A. Wagner and members of our lab for comments on the manuscript.

Correspondence and requests for materials should be addressed to N.K. (e-mail: ngk@psyche.mit.edu).

Involvement of visual cortex in tactile discrimination of orientation

Andro Zangaladze*, Charles M. Epstein, Scott T. Grafton & K. Sathian

Department of Neurology, Emory University School of Medicine, WMRB-6000, PO Drawer V, 1639 Pierce Drive, Atlanta, Georgia 30322, USA

The primary sense modalities (vision, touch and so on) are generally thought of as distinct. However, visual imagery is implicated in the normal tactile perception of some object properties, such as orientation¹, shape and size². Furthermore, certain tactile tasks, such as discrimination of grating orientation¹ and object recognition³, are associated with activity in areas of visual cortex. Here we show that disrupting function of the occipital cortex using focal transcranial magnetic stimulation (TMS) interferes with the tactile discrimination of grating orientation. The specificity of this effect is illustrated by its time course and spatial restriction over the scalp, and by the failure of occipital TMS to affect either detection of an electrical stimulus applied to the fingerpad or tactile discrimination of grating texture. In contrast, TMS over the somatosensory cortex blocked discrimination of grating texture as well as orientation. We also report that, during

* Present address: Department of Neurology, Thomas Jefferson University Hospital, 1025 Walnut Street, Suite 310, Philadelphia, Pennsylvania 19107, USA.

tactile discrimination of grating orientation, an evoked potential is recorded over posterior scalp regions with a latency corresponding to the peak of the TMS interference effect (about 180 ms). The findings indicate that visual cortex is closely involved in tactile discrimination of orientation. To our knowledge, this is the first demonstration that visual cortical processing is necessary for normal tactile perception.

Visual cortex is known to be involved in nonvisual perception in blind humans⁴⁻⁹. This has been attributed to neural plasticity resulting from visual deprivation. However, visual processing can also influence certain aspects of normal human tactile perception. For instance, during haptic sorting, instructions to emphasize visual images bias subjects to sort by object properties such as size and shape, whereas without such instructions sorting is based on texture and hardness². Using positron emission tomography (PET), we showed that discrimination of the orientation of a grating on the fingerpad is associated with subjective reports of visual imagery and increased regional cerebral blood flow, relative to that seen during discrimination of grating texture (spacing), in a contralateral region of extrastriate visual cortex near the parieto-occipital fissure¹. This locus, which is also active during visual discrimination of grating orientation¹⁰ and other tasks requiring spatial mental imagery¹¹, may be the human homologue of an area in the parieto-occipital fissure of macaques (V6/PO) where a large proportion of neurons are orientation-selective¹². A functional magnetic resonance imaging (fMRI) study found that visual cortical areas are also active during tactile object recognition, compared with texture discrimination³. The recruitment of processing in visual cortex may reflect top-down activation of visual representations to facilitate tactile discrimination of orientation or shape. Alternatively, the observed activity could be merely an epiphenomenon.

One way to distinguish between these two possibilities is to interfere with processing in visual cortex. To do this, we applied TMS over the occipital scalp, a procedure that blocks visual perception by disrupting the function of extrastriate visual cortex^{13,14}. In Experiment 1, TMS pulses were delivered to multiple sites over occipital cortex while subjects attempted to discriminate the orientation of a grating on the right index fingerpad (Fig. 1) with their eyes closed. At each site, TMS was applied 10, 180 or 400 ms after the onset of grating presentation. To control for nonspecific effects of the loud noise accompanying TMS, performance was also measured with the magnetic stimulator in air.

Figure 2a shows a clear effect of TMS over occipital cortex at the higher of the two midline sites tested (M4) and contralateral to the hand encountering the grating (L3). Compared with discrimination performance with the stimulator in air, performance at these sites

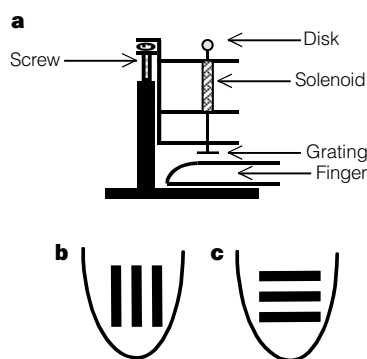


Figure 1 Set up used for grating application. **a**, The screw enabled positioning of the grating just above the fingerpad by a dovetail arrangement. From this position, activation of the solenoid caused downward displacement of the grating by about 3 mm. Rotating the disk changed grating orientation, which was either along (**b**) or across (**c**) the long axis of the finger.

was unaffected by TMS at the 10-ms delay but was markedly impaired at the 180-ms delay. By 400 ms, the effect was slight. In contrast, TMS ipsilateral to the hand feeling the grating (R3) and at the lower midline site (M2) had no effect (Fig. 2a). This spatial restriction, together with the clear time course of the effect, indicates that the impairment of perception was not due to muscle contraction or other nonspecific effects of TMS. Performance was significantly different between the 10-ms and 180-ms delays at the M4 ($t = 11.58$, d.f. = 10, $P = 0.00000004$) and L3 ($t = 3.79$, d.f. = 9, $P = 0.004$) sites but not at the R3 ($t = 0.25$, d.f. = 9, $P = 0.81$) or M2 ($t = 2.14$, d.f. = 4, $P = 0.1$) sites.

To test whether the disruption of tactile performance by occipital TMS was specific to discrimination of orientation or represented a more general effect on somatosensory perception, we investigated the effect of occipital TMS on a control task, detection of a suprathreshold electrical stimulus to the fingerpad. Figure 2b shows that performance on this task was unaffected. There were no significant performance differences between the 10-ms and 180-ms delays for TMS at the M4 ($t = -1.83$, d.f. = 4, $P = 0.14$), L3 ($t = -1$, d.f. = 4, $P = 0.37$) or R3 ($t = -1.29$, d.f. = 4, $P = 0.27$) sites. The lack of effect of TMS on this control task may indicate that the effect of TMS is specific to orientation discrimination, rather than being a general suppression of somatosensory processing. However, the control task was slightly easier than the orientation task (performance with the stimulator in air was 94% correct versus 85% correct; Fig. 2b, a) and was also a simple detection task, whereas the main task involved discrimination (of grating orientation). Hence, we conducted another experiment (Experiment 2), in which the control task was discrimination of grating texture (spacing task), the same control task as in our previous PET study¹. As the locus of activation (orientation minus spacing) in the PET study¹ was superior and lateral to the M4 site, we applied TMS pulses directly over the PET activation locus (L6) in the left hemisphere and the homologous area in the right hemisphere (R6), as well as to M4, using only the 10- and 180-ms delays at these sites in Experiment 2 (as there was little effect at 400 ms in Experiment 1). We also delivered TMS to left primary somatosensory cortex, using a 30-ms delay. Subjects were blindfolded in Experiment 2.

In Experiment 2, TMS at M4 clearly interfered with discrimination of grating orientation at 180 ms but not at 10 ms (Fig. 2c), replicating the effect obtained in Experiment 1. A similar effect was obtained at L6, the PET activation locus, but not at R6, its

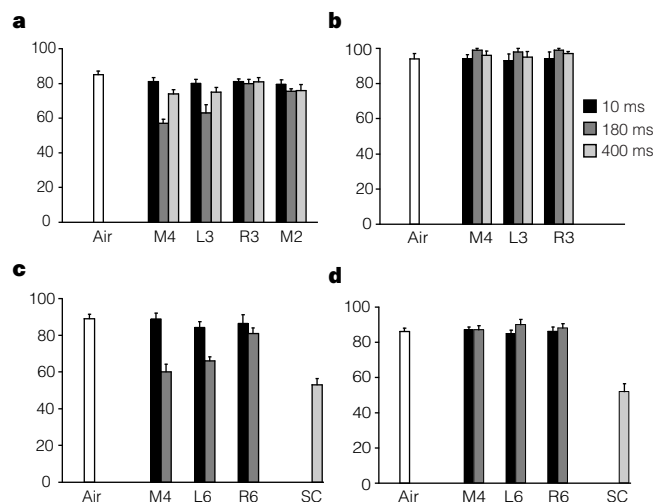


Figure 2 Effect of TMS at the delays indicated. Bar graphs show mean performance (% correct responses) and standard errors. **a**, Experiment 1, orientation task; **b**, Experiment 1, electrical stimulation; **c**, Experiment 2, orientation task; **d**, Experiment 2, spacing task. SC, TMS over somatosensory cortex (30-ms delay). Air, stimulator in air.

homologue in the opposite hemisphere (Fig. 2c). Performance was significantly poorer at the 180-ms delay than at the 10-ms delay at M4 ($t = 6.75$, d.f. = 5, $P = 0.001$) and L6 ($t = 4.22$, d.f. = 5, $P = 0.008$) but not R6 ($t = 2.22$, d.f. = 5, $P = 0.08$). In contrast, occipital TMS had no effect on discrimination performance in the spacing task (Fig. 2d). There were no significant performance differences between the 10- and 180-ms delays at any of the sites (M4: $t = 0$, d.f. = 5, $P = 1$; L6: $t = -1.46$, d.f. = 5, $P = 0.2$; R6: $t = -0.39$, d.f. = 5, $P = 0.71$). Notably, baseline performances with the stimulator in air were almost identical in the orientation (89% correct; Fig. 2c) and spacing (86% correct; Fig. 2d) tasks. Thus, the effect of occipital TMS was specific to the orientation task. Consistent with this, most subjects felt that mental visualization of the grating was helpful in the orientation task but not the spacing task.

Stimulation over primary somatosensory cortex, however, suppressed discrimination performance to near-chance levels in both orientation and spacing tasks (Fig. 2c, d). The effects were statistically significant: performance was poorer in the TMS condition than with stimulation in air (orientation: $t = 5.55$, d.f. = 2, $P = 0.03$; spacing: $t = 20$, d.f. = 2, $P = 0.002$). Thus, in contrast to the selective effect of occipital TMS on the orientation task, TMS over somatosensory cortex was nonselective, affecting both orientation and spacing discriminations equally. Subjective reports also differed according to where TMS was applied. At occipital sites where TMS impaired performance on orientation discrimination, subjects reported that they could still feel the grating but were unsure of its orientation. In contrast, with TMS over primary somatosensory cortex, subjects reported interference with perception of the grating itself, often feeling only a sensation of pressure but being unable to distinguish the pattern of the grating.

Figure 3 shows that our results are consistent with the idea that impairment of tactile discrimination of grating orientation by occipital TMS is due to stimulation of the region of parieto-occipital cortex identified by PET¹. Stimulation at effective sites (M4, L3 and

L6) induced electric fields whose calculated magnitudes were consistent with physiological effects at the PET activation locus. In contrast, the fields derived from stimulation at ineffective sites (R6, R3 and M2) failed to enclose the area of PET activation. Although the cerebellum has been implicated in somatosensory discrimination¹⁵, it is unlikely that our results were due to cerebellar stimulation, as there was no effect at the most inferior site (M2), which was the closest site to the cerebellum.

We also studied the distribution and time course of the averaged electrical potential evoked over the scalp in response to stimulation of the fingerpad with the grating. To isolate neural processing that was specific to discrimination of orientation, the potentials evoked when subjects were engaged in this task were compared with those obtained while subjects performed a control task, counting the number of tactile stimuli. This controlled for attentive somatosensory processing. The resulting subtraction potential (Fig. 4) showed a peak at 150–180 ms in contralateral parietal and occipital leads, consistent with the time course of the TMS interference effect.

Together with the subjective reports of visual imagery in this task and the associated parieto-occipital cortical activation noted previously¹, our findings support the proposal that visual processing facilitates normal tactile discrimination of orientation. Visual processing may similarly facilitate tactile object recognition³. Perhaps this is related to the fact that we generally rely on the visual system for orientation and shape discrimination. We emphasize that this dependence on the visual system does not apply to all aspects of tactile perception, as shown by the failure of occipital TMS to affect the detectability of an electrical stimulus applied to the fingerpad or discrimination performance on the spacing task. Thus, involvement of visual cortex may be beneficial when macrogeometric features such as orientation and shape are to be discriminated, but not for microgeometric features such as texture. In an earlier study⁵, occipital TMS affected the ability of blind but not sighted subjects to identify embossed Roman letters. The lack of effect in the sighted may reflect the dependence of this task, like texture discrimination, on processing of microspatial detail¹⁶. Alternatively, the different set sizes used for sighted (5-letter sets) and blind subjects (26-letter sets) may underlie the differential effects in these two subject groups, given the intrinsic confusability of certain tactually presented letters¹⁷—it remains possible that sighted subjects could also show the effect under different experimental conditions. Our findings provide the first demonstration, to our knowledge, that visual cortex is necessary for optimal tactile performance in normally sighted subjects. □

Methods

Subjects

Fourteen subjects volunteered after giving informed consent and were paid for their participation. All were right-handed. None had any history of neurological disease, trauma to the hands or their innervation, or fingerpad calluses.

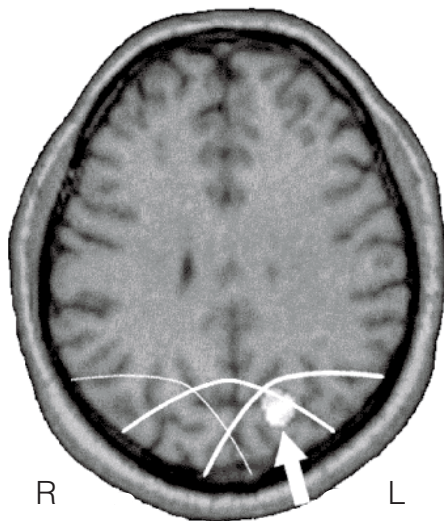


Figure 3 Activation due to grating orientation (relative to spacing) observed in previous PET study¹ superimposed on an axial, T1-weighted MRI slice derived from one subject who participated in both the PET study and the present study. Arrow, activation. Talairach coordinates of centre of activation: $x = -20$, $y = -71$, $z = 32$. Isopotential lines of the TMS-induced electric fields corresponding to 90% of the resting motor threshold are also superimposed on the MRI. Brain regions on the concave side of the curves are exposed to electric fields above the threshold of known physiological effects. The heavy curves refer to TMS at the (effective) M4 and L3 sites; the thin curve refers to the (ineffective) R3 site. Similar calculations (data not shown) indicated the compatibility of the effectiveness of TMS at other occipital sites (M2, L6, R6) in the orientation task with the ability to evoke physiological effects at the PET activation locus.

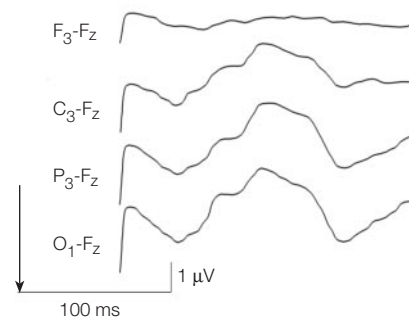


Figure 4 Electrical potential (averaged across subjects) recorded over the scalp during tactile discrimination of grating orientation, relative to counting of tactile stimuli. Arrow, onset of grating contact with fingerpad.

Procedures

Tactile discrimination tasks. Plastic, steel-backed gratings measuring about 20 mm² and with equal groove and ridge widths were manufactured as described¹⁸. An electromechanical device was used to apply the gratings (for 63-ms duration) to the index fingerpad of the right hand, oriented either along or across the long axis of the finger (Fig. 1). The ability to discriminate orientation increases monotonically as grating groove width (GW) increases up to 3 mm (ref. 19). Subjects were first tested psychophysically on each task. For the orientation task, the discrimination threshold was taken as the GW yielding 75% correct performance²⁰. Threshold determinations were based on at least 20 trials for each grating. A grating yielding performance that was above threshold but below ceiling was then selected for each subject. This grating had the maximal GW of 3 mm for most subjects and 2 mm for others. For the spacing task, a similar procedure was followed, resulting in selection of a pair of gratings for each subject (GW 3 mm and 2, 1.5 or 1.2 mm). Gratings were applied in either orientation in the spacing task, at random. Both tasks used two-alternative forced-choices, subjects responding whether the grating was oriented along or across the finger (in the orientation task) or whether the grating grooves/ridges were wide or narrow (in the spacing task). Performance was expressed as the percentage of correct responses.

Electrical stimulation. Electrical stimulation was delivered to the right index fingerpad of five subjects in Experiment 1. A suprathreshold stimulus intensity was chosen for each subject.

TMS. TMS was applied using a Cadwell MES-10 (Cadwell Laboratories) stimulator with a custom high-efficiency iron-core coil that measured 12 cm × 7.5 cm. This coil induces an electric field similar to a figure-of-eight coil, with its maximum directly below the centre²¹. Stimuli were cosine pulses of 200 µs duration, at 150% of the relaxed motor threshold²². The inter-stimulus interval was at least 1 s, to avoid seizures. An electronic circuit incorporating a variable delay was used to deliver TMS at particular delays following the onset of the tactile stimulus, registered electrically in the case of gratings with the help of silver gel. On the basis of pilot studies indicating that the peak of the TMS interference effect over occipital cortex occurred at 150–200 ms, the 180-ms delay was chosen as representative of the peak effect. For TMS in air, the coil was held close to but not in contact with the scalp and rotated to direct the magnetic field away from the head.

In Experiment 1, 11 subjects were tested with TMS applied at the M4 (occipital midline, 4 cm above theinion), L3 and R3 sites (both 3 cm above and 4 cm lateral to theinion on the left and right, respectively), except for one subject who dropped out after testing at M4 at the 10- and 180-ms delays. TMS was applied at the M2 site (2 cm above theinion in the midline) in 5 of the 11 subjects. Six subjects (including three who had taken part in Experiment 1) were tested with occipital TMS in Experiment 2, at L6, R6 (both 6 cm above and 2 cm lateral to theinion on the left and right, respectively) and M4. For stimulation over primary somatosensory cortex, the site over the left hemisphere from which a motor response in the right thumb and/or index finger could just be evoked was located. The coil was then moved posteriorly until the motor response just disappeared, and this site was chosen as the somatosensory cortical site. In two out of three subjects, this localization was aided by reports of somatic sensations in the hand—warmth in one subject and numbness in another. As the early peak (presumably from primary somatosensory cortex) in the evoked potential response to stimulation with the grating occurred at 30–50 ms (data not shown), a delay of 30 ms was chosen for TMS at this site. Owing to the discomfort resulting from stimulation at this site (due to muscle contraction), only a single delay was used and only three subjects were tested. The stimulus intensity at this site had to be reduced to the motor threshold for two subjects as they could not tolerate stimulation at higher intensities. Trials were presented in interleaved blocks of 10 with a minimum of 20 trials at each delay and location.

To determine the relationship between the sites at which effects of occipital TMS were obtained and the previously reported PET activation locus¹, we used the averaged resting motor threshold²² in our subjects to scale the TMS-induced electric fields in a model head^{13,23}. The approximate boundaries for physiological effects of TMS (taken as 90% of the resting motor threshold²⁴) at selected sites are displayed in Fig. 3.

Statistical testing. Paired *t*-tests (two-tailed; $\alpha = 0.05$) were used to assess the statistical significance of differences between conditions.

Evoked potentials. Evoked potential recordings were carried out in five subjects in response to tactile stimulation of the right index fingerpad with a grating, as described above. The evoked potentials were averaged with band-pass filters of 0.1–200 Hz. In one session, the active scalp electrodes were positioned on the left side of the head over frontal (F₃), central (C₃), parietal (P₃) and occipital (O₁) sites. In another session, recordings were made with the active scalp electrodes in the corresponding positions on the right. A midline frontal electrode (F_z) was used as the reference. In each session, two runs of 100–120 trials were performed for each condition (discrimination of grating orientation and counting stimuli). The two conditions alternated with each other and the order was counterbalanced across subjects. To ensure that the posterior location of the observed peak in Fig. 4 was not an artefact of the frontal reference electrode used, evoked potentials were also recorded in two subjects using a linked earlobe reference electrode. The potential thus obtained (data not shown) was similar in scalp distribution and time course to that in Fig. 4.

Received 14 July; accepted 2 August 1999.

- Sathian, K., Zangaladze, A., Hoffman, J. M. & Grafton, S. T. Feeling with the mind's eye. *NeuroReport* **8**, 3877–3881 (1997).
- Klatzky, R. L., Lederman, S. & Reed, C. There's more to touch than meets the eye: the salience of object attributes for haptics with and without vision. *J. Exp. Psychol. Gen.* **116**, 356–369 (1987).
- Deibert, E., Kraut, M., Kremen, S. & Hart, J. Neural pathways in tactile object recognition. *Neurology* **52**, 1413–1417 (1999).

- Sadato, N. *et al.* Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* **380**, 526–528 (1996).
- Cohen, L. G. *et al.* Functional relevance of cross-modal plasticity in blind humans. *Nature* **389**, 180–183 (1997).
- Sadato, N. *et al.* Neural networks for Braille reading by the blind. *Brain* **121**, 1213–1229 (1998).
- Büchel, C., Price, C., Frackowiak, R. S. J. & Friston, K. Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain* **121**, 409–419 (1998).
- Röder, B., Rösler, F., Henninghausen, E. & Näcker, F. Event-related potentials during auditory and somatosensory discrimination in sighted and blind subjects. *Cognit. Brain Res.* **4**, 77–93 (1996).
- Kujala, T. *et al.* Visual cortex activation in blind humans during sound discrimination. *Neurosci. Lett.* **183**, 143–146 (1995).
- Sergent, J., Ohta, S. & MacDonald, B. Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain* **115**, 15–36 (1992).
- Mellet, E. *et al.* Functional anatomy of spatial mental imagery generated from verbal instructions. *J. Neurosci.* **16**, 6504–6512 (1996).
- Galletti, C., Battaglini, P. P. & Fattori, P. Functional properties of neurons in the anterior bank of the parieto-occipital sulcus of the macaque monkey. *Eur. J. Neurosci.* **3**, 452–461 (1991).
- Epstein, C. M. & Zangaladze, A. Magnetic coil suppression of extrafoveal visual perception using disappearance targets. *J. Clin. Neurophysiol.* **13**, 242–246 (1996).
- Epstein, C. M., Verson, R. & Zangaladze, A. Magnetic coil suppression of visual perception at an extracalcarine site. *J. Clin. Neurophysiol.* **13**, 247–252 (1996).
- Gao, J.-H. *et al.* Cerebellum implicated in sensory acquisition and discrimination rather than motor control. *Science* **272**, 545–547 (1996).
- Johnson, K. O. & Hsiao, S. S. Neural mechanisms of tactual form and texture perception. *Annu. Rev. Neurosci.* **15**, 227–250 (1992).
- Vega-Bermudez, F., Johnson, K. O. & Hsiao, S. S. Human tactile pattern recognition: active versus passive touch, velocity effects, and patterns of confusion. *J. Neurophysiol.* **65**, 531–546 (1991).
- Sathian, K. & Zangaladze, A. Tactile learning is task-specific but transfers between fingers. *Percept. Psychophys.* **59**, 119–128 (1997).
- van Boven, R. W. & Johnson, K. O. The limit of tactile spatial resolution in humans: grating orientation discrimination at the lip, tongue and finger. *Neurology* **44**, 2361–2366 (1994).
- Sathian, K. & Zangaladze, A. Tactile spatial acuity at the human fingertip and lip: Bilateral symmetry and inter-digit variability. *Neurology* **46**, 1464–1466 (1996).
- Epstein, C. M. *et al.* Optimum stimulus parameters for lateralized suppression of speech with magnetic brain stimulation. *Neurology* **47**, 1590–1593 (1996).
- Pascual-Leone, A. *et al.* Safety of rapid-rate transcranial magnetic stimulation in normal volunteers. *EEG Clin. Neurophysiol.* **89**, 120–130 (1993).
- Epstein, C. M., Schwartzberg, D. G., Davey, K. R. & Sudderth, D. B. Localizing the site of magnetic brain stimulation in humans. *Neurology* **40**, 666–670 (1990).
- Ziemann, U., Rothwell, J. C. & Ridding, M. C. Interaction between intracortical inhibition and facilitation in human motor cortex. *J. Physiol. (Lond.)* **496**, 873–881 (1996).

Acknowledgements

This work was supported by grants from the National Institute of Neurological Disorders and Stroke and the National Eye Institute to K.S. We thank D. Rice for machining and electronic support.

Correspondence and requests for materials should be addressed to K.S. (e-mail: ksathia@emory.edu).

Primate spinal interneurons show pre-movement instructed delay activity

Yifat Prut & Eberhard E. Fetz

Department of Physiology & Biophysics and the Regional Primate Research Center, Box 357330, University of Washington, Seattle, Washington 98195, USA

Preparatory changes in neural activity before the execution of a movement have been documented in tasks that involve an instructed delay period (an interval between a transient instruction cue and a subsequently triggered movement). Such preparatory activity occurs in many motor centres in the brain, including the primary motor cortex^{1–6}, premotor cortex^{7–9}, supplementary motor area^{6,10,11} and basal ganglia^{6,12,13}. Activity during the instructed delay period reflects movement planning, as it correlates with parameters of the cue and the subsequent movement (such as direction and extent^{5,6,9}), although it occurs well before muscle activity. How such delay-period activity shapes the ensu-

ing motor action remains unknown. Here we show that spinal interneurons also exhibit early pre-movement delay activity that often differs from their responses during the subsequent muscle activity. This delay activity resembles the set-related activity found in various supraspinal areas, indicating that movement preparation may occur simultaneously over widely distributed regions, including spinal levels. Our results also suggest that two processes occur in the spinal circuitry during this delay period: the motor network is primed with rate changes in the same direction as subsequent movement-related activity; and a superimposed global inhibition suppresses the expression of this activity in muscles.

The generation of voluntary movements is commonly thought to develop sequentially in the motor system^{3,9}, beginning with preparatory activity in higher motor areas and propagating through the premotor and primary motor cortex (M1) to the spinal cord. Alternatively, these motor areas may operate in a predominantly parallel mode^{6,12}, in which movement-related activity would effectively appear simultaneously in many motor areas, regardless of their relative synaptic 'distance' from muscles⁶. Indeed, indirect evidence has indicated that early stages of motor preparation may involve circuits as peripheral as the spinal cord. Spinal reflexes, both mono- and polysynaptic, can be modulated during an instructed delay period^{15–18}. This modulation is not mediated through the γ -motor-neuronal system^{19,20} and has been partially attributed to presynaptic inhibition¹⁵. Another contribution to reflex modulation may arise from spinal interneurons (INs), which receive direct inputs from M1 and higher premotor areas^{14,21–24}. The role of

Table 1 Proportions of spinal INs showing significant delay modulation

	T-test		Sign rank	
	No.	%	No.	%
Total SDM	133	33.5*	149	37.5*
SDM in pre-extension only	59	44.4†	62	41.6†
SDM in pre-flexion only	40	30.1†	45	30.2†
SDM in both pre-extension and pre-flexion‡	34***	25.6†	42***	28.2†

* Per cent of the total number of valid cells ($n = 397$).

† Per cent of the number of SDM cells ($n = 133$).

‡ For cells with SDM in both flexion and extension, the expected number of cases was computed assuming independence of SDM occurrence in flexion and extension trials. In both cases the counts were significantly higher than expected (Poisson, $P < 0.001$ ***).

segmental INs in preparatory changes occurring before voluntary movement is unknown.

To address this issue we investigated the activity of cervical INs in monkeys performing an isometric flexion–extension instructed delay task (Fig. 1). Such a task separates motor preparation from execution of movement. Most INs in awake monkeys are active in the absence of active torque, unlike motor neurons²⁵. We searched for cells that modulated their activity during a post-instruction delay period, relative to their activity during rest.

Figure 2 shows the activity of three INs whose firing rate was modulated during the instructed delay period. They show increased activity during one or both delay periods (Fig. 2b, d) and suppression during both delay periods (Fig. 2c). For comparison, Fig. 2a shows an IN that is activated during active torque but not during the delay period. Of 450 cells recorded from two monkeys, 397 had background activity at rest and were analysed for the existence of delay-related changes in activity (Table 1). About one-third of these INs showed significant delay modulation (SDM) in flexion and/or extension trials (paired t -test, $P < 0.05$). The median onset time of SDM after the onset of the cue was 190 ms (mean 227.3 ms). Inhibitory SDMs tended to begin earlier than excitatory SDMs (median, 170 and 230 ms, respectively). The proportions of set-related and movement-related cells in the spinal cord were similar to the proportions found in other motor areas (Table 2). The main difference was the relatively small number of 'pure' set-related spinal INs, namely INs with no movement-related but only set-related activity.

An obvious possibility is that the SDM could have been associated with low levels of muscle activity. Electromyogram (EMG) activity of forearm muscles was routinely recorded in parallel with INs and revealed no muscle activity in either flexors or extensors during the delay (Fig. 1). Detailed analysis of data from 17 recording sites (from which 69 cells were recorded) showed no evidence of significant muscle activity in the delay periods.

Delay activity in spinal INs could also represent the early onset of subthreshold movement-related rate modulation. In this case, the change in firing during the delay period would have the same polarity (excitation or inhibition) as the subsequent modulation during active torque. This, however, was not the general case. For the 167 cases of SDM (pre-flexion or pre-extension), Fig. 3 shows the average change in rate during the delay period relative to the preceding rest period (abscissa) plotted against the corresponding rate modulation during the hold period (ordinate). Sixty-four per cent of the points lie in the two quadrants of the graph corresponding to the 'congruent' case, in which the SDM changes were in the same direction as the change of rate during the active torque period (lower-left and upper-right quadrants). Thirty-four per cent of the points fall in the upper-left quadrant, representing a decrease in rate during the delay followed by a rate increase during the active torque. INs tended to be inhibited more than excited during the delay period compared with the active torque period. The proportions of cells with inhibition versus excitation in the delay period were significantly different from the proportions found for the hold-period activity (McNemar's test, $P < 0.01$). Suppression is also common when inhibition is seen during the delay period

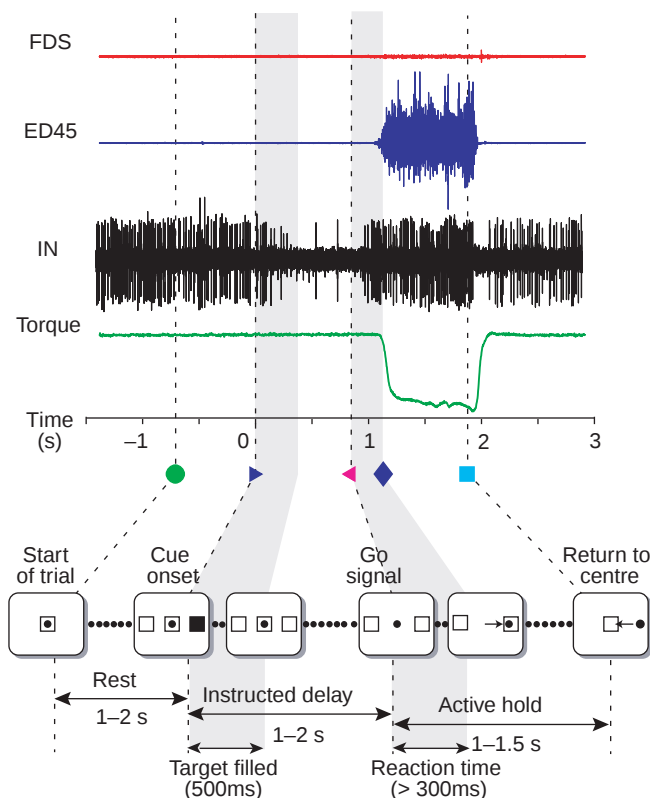


Figure 1 Example of a single instructed delay trial. Traces (from top) show activity of two muscles (flexor digitorum sublimis (FDS, red) and extensor digitorum 4 and 5 (ED45, blue)), activity of IN (black) and torque signal (green). Bottom, sequence and timing of events during a single trial (see Methods for details). During the delay period, the firing rate of this unit decreases, with no accompanying torque deflection or EMG activity of either muscle.

in other motor structures including M1 (ref. 4), basal ganglia¹³ and premotor cortex²⁶. In particular, changes in primary motor cortex neurons were similar to those in spinal INs, although fewer M1 cells had opposite responses during delay and movement (16% compared with 34% in our study). Most of these reversals in M1 involved inhibition during the delay and activation during movement, similar to our INs.

Parametric studies showed that cortical delay activity could be related to several parameters of the subsequent movement, including the direction^{1,5,6,26} and extent^{5,9}. We tested the coding of these parameters by spinal INs. The directional selectivity of each IN during the instructed delay period was evaluated by comparing its firing rate during the delay period in flexion trials with its rate during the same period in extension trials. Out of 450 cells, 43 (10%) exhibited directionality in firing during the delay period (analysis of variance (ANOVA), $P < 0.05$), 166 (37%) during the pre-active period (from go signal until torque onset) and 207 (46%) during the active torque period. Seventeen of the 43 cells (40%)

were directional during the delay but not during the active torque period. Most INs had the same polarity of delay activity in both flexion and extension trials, as opposed to reciprocal activation. Similar results were obtained for the relation between torque magnitude and delay period activity of INs.

The functional identity of the cells that showed SDM is of considerable interest. Chronic recording conditions limit our ability to identify the types of IN recorded in terms of reflex properties²⁵. However, INs with functional linkages to the recorded forearm muscles could be identified by spike-related changes in EMG activity²⁵. Of 300 tested INs, 46 had functional linkages with muscle activity. Those INs were significantly ($P < 0.01$) more likely to exhibit SDM (27/46, 59%) than INs without such links (94/254, 37%). In both groups SDM was most often inhibitory.

A comparison of activity changes during the instructed delay with activity during the subsequent movement period indicates that there may be two types of delay-period activity in the spinal cord. Some INs showed a 'congruent' pattern in which the delay activity

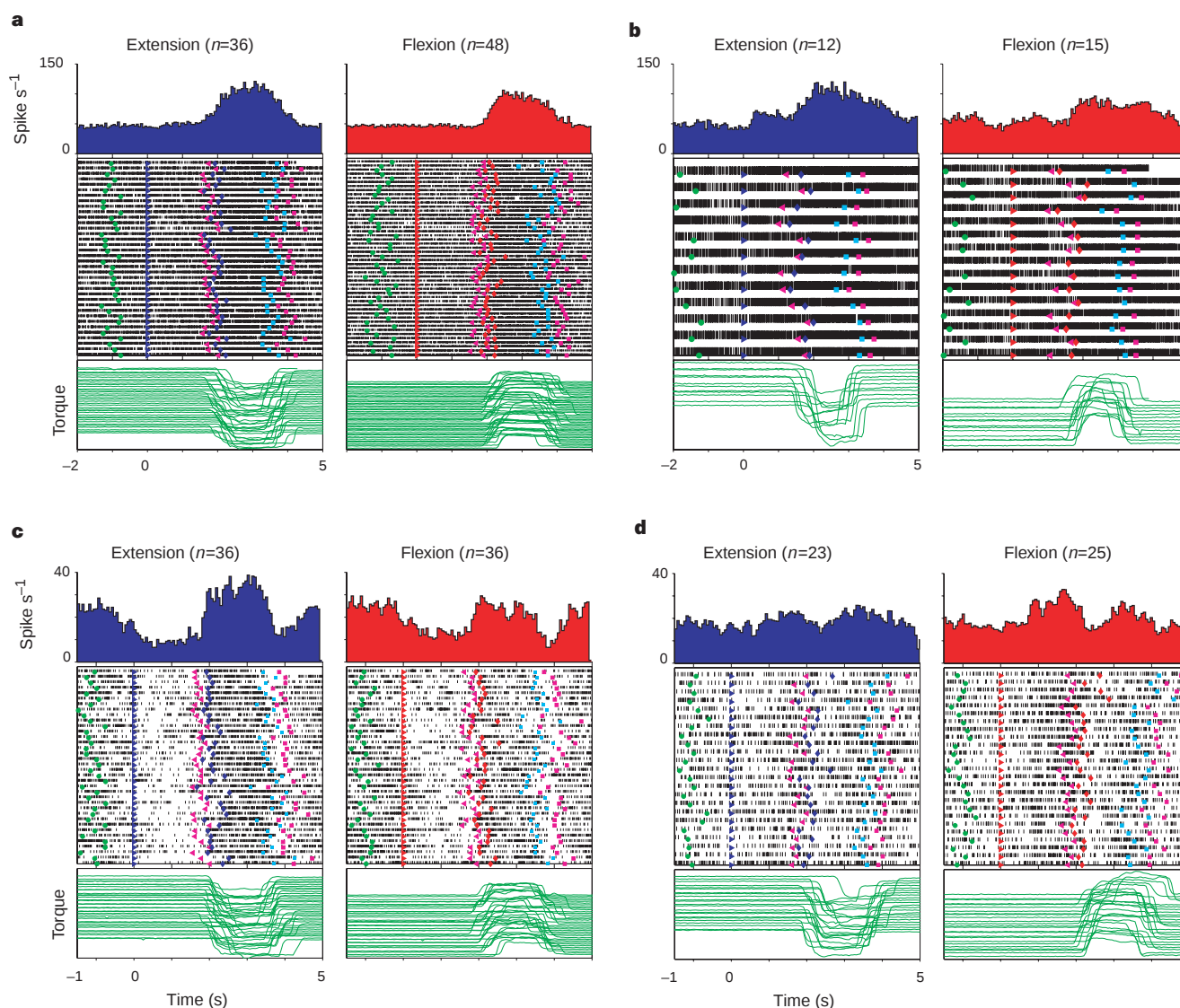


Figure 2 Response properties of spinal INs. Trials are aligned with cue onset (blue and red triangles). Lowest traces (green) show the single trial torques. The raster plots present the unit activity. The symbols code trial onset (green circles), go signal (pink triangles), movement onset (red and blue diamonds), return to centre signal (cyan squares) and movement offset (pink squares). The peri-stimulus time histograms (top) quantify the IN's

stimulus-locked rate modulations. **a**, IN with no rate modulation during delay but increased rate during active torque periods. **b**, IN with a pre-flexion excitatory delay modulation. **c**, IN with inhibitory delay modulations in flexion and extension trials. **d**, IN with excitatory delay modulations in flexion trials.

Table 2 Response properties of neurons in different motor areas

Area	Spinal cord	Primary motor (M1)	Pre-motor	SMA	Putamen
Ref. no.	6	10	6	6	12
Set only (%)	2.5	10.9	20.9	32.5	24.3
Movement only (%)	67.4	62.9	50.6	45.5	66.9
Set + Movement (%)	30.1	26.2	28.6	22.5	8.8
n	405	202	91*	222	317

Relative proportions of set-related, movement-related, and set + movement-related cells found in different regions of the motor system. Note that different studies used different statistical methods to determine response properties of the cells.

* From these data we omitted cells with a pure signal-related response (8 cells).

had the same polarity (albeit with smaller amplitude) as the modulation during active torque (lower-left and upper-right quadrants of Fig. 3). This type of SDM activity appeared to be locked in time to cue onset, and had a relatively late onset time. The second type of delay-period activity was largely inhibitory during the delay, irrespective of the subsequent active-torque rate modulation (left quadrants of Fig. 3). This inhibitory activity seemed to be more loosely locked to the cue onset and to begin earlier than excitatory SDMs. The two quadrants that contain the 'purest' examples of these two groups (upper-left and upper-right quadrants) had the greatest difference in onset times: median 250 ms (mean 257.2 ms) versus 110 ms (mean 172.3 ms). This difference was significant ($P < 0.01$, Mann-Whitney test). The inhibitory cases in the lower left quadrant could represent either mechanism and had SDMs with onset times between the two extreme values. The congruent pattern can be understood as a subthreshold preparation for movement involving the priming of cells in the direction in which they must fire during movement execution. The inhibitory delay patterns could reflected a superimposed general 'braking' mechanism, which suppresses the tendency to initiate movement. This inhibition is released at the onset of the go signal, when movement starts. The fact that many cells with functional linkages to muscle activity show inhibitory SDM supports this interpretation.

The source of the SDM is probably supraspinal, as afferent input to the cord is the same during rest and the delay period^{19,20}. An obvious candidate is the cerebral cortex, because arm-related premotor areas exhibit robust delay activity and project to the spinal cord^{14,24} (mostly to the intermediate grey zone—layers V–VII

(ref. 21). Indeed, it has been suggested that the premotor cortex inhibits motor pathways^{27,28}, as has been shown during epileptic activity²⁹.

Previous evidence indicated that supraspinal motor centres operate in a distributed parallel mode^{3,6,14}. Our results support and extend this view to spinal levels, where segmental INs exhibit instructed delay activity whose timing and properties are similar to those observed in supraspinal centres. This contrasts with the traditional view that the spinal cord is simply an output relay, executing motor plans that are pre-generated in supraspinal structures. It indicates that motor cortical areas may interact continuously with spinal networks during the earliest stages of preparation for movement. The modulation of spinal INs during the instructed delay could not only prepare for the subsequent motor actions but also modulate transmission of afferent information from the somatic periphery to spinal¹ and supraspinal levels. □

Methods

Animals and behavioural task.

Two monkeys (*Macaca nemestrina*) performed a flexion–extension task with a delay. The monkey controlled a cursor on a computer screen (black circle in Fig. 1) by isometric torque. A trial was initiated by the appearance of a central box. The monkey positioned the cursor inside the box by generating zero torque for a rest period of 1–2 s. Then two symmetric targets appeared in the flexion and extension positions with one target filled for 500 ms (cue), defining the onset of a delay period. The disappearance of the central box, 1–2 s after cue onset, served as the go signal. The monkey then had to acquire the previously filled target by generating an isometric torque in the appropriate direction, and to keep the cursor within the target box for an active-torque period of 1–1.5 s. After this period the two target boxes disappeared and the central box reappeared. The monkey returned to the rest position and received a reward, after which the screen went blank for 500 ms and a new trial started.

Recording sessions.

Details of the recording technique are described elsewhere²⁵. A chamber was implanted above the cervical spinal cord (C_6 – T_1). We recorded single-unit activity with tungsten electrodes and EMG activity from 6–12 forearm muscles.

Data analysis.

We selected cells with a firing rate above a minimal level (median of 1.5 spikes per s) in the rest and/or delay period. Rest period was the time from trial onset to cue onset. The delay period was defined as the time from cue onset until either onset of the go signal or 300 ms before movement onset, whichever occurred first. SDM was measured by comparing the single trial rates during the delay period with the corresponding rates during the rest period. This test identifies consistent rate differences across trials. The results were then confirmed using a nonparametric sign-rank (Wilcoxon) paired test.

Received 3 March; accepted 23 July 1999.

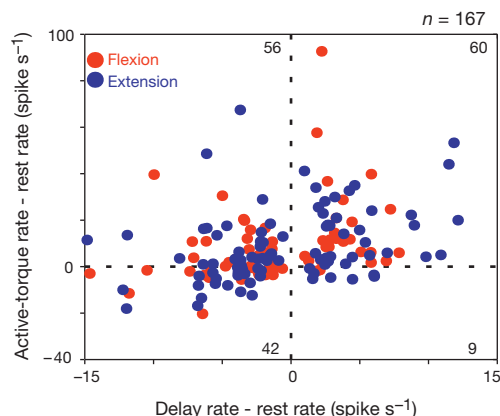


Figure 3 Relations between the magnitude of the SDM and the modulation in firing during active torque. x axis plots the rate modulation during delay period relative to rest. The values were computed by averaging for each cell the single-trial difference in firing rate during the delay period and the rest period. y axis values are the rate modulations relative to rest during active torque period. Each dot represents a cell with an SDM during pre-flexion (black) and/or pre-extension (grey) periods (34 cells contributed values for both flexion and extension). The numbers of points in each quadrant are given in the corners.

1. Tanji, J. & Evarts, E. V. Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. *J. Neurophysiol.* **39**, 1062–1068 (1976).
2. Kubota, K. & Hamada, I. Preparatory activity of monkey pyramidal tract neurons related to quick movement onset during visual tracking performance. *Brain Res.* **168**, 435–439 (1979).
3. Thach, W. T. Correlation of neural discharge with pattern and force of muscular activity, joint position and direction of intended next movement in motor cortex and cerebellum. *J. Neurophysiol.* **41**, 654–676 (1978).
4. Georgopoulos, A. P., Crutcher, M. D. & Schwartz, A. B. Cognitive motor processes. III. Motor cortical prediction of movement direction during an instructed delay period. *Exp. Brain Res.* **75**, 183–194 (1989).
5. Riehle, A. & Requin, J. Neuronal correlates of the specification of movement direction and force in four cortical areas of the monkey. *Behav. Brain Res.* **70**, 1–13 (1995).
6. Alexander, G. E. & Crutcher, M. D. Preparation for movement: neural representations of intended direction in three motor areas of the monkey. *J. Neurophysiol.* **64**, 133–150 (1990).
7. Weinrich, M. & Wise, S. P. The premotor cortex of the monkey. *J. Neurosci.* **2**, 1329–1345 (1982).
8. Wise, S. P., Weinrich, M. & Mauritz, K. H. Movement-related activity in the premotor cortex of rhesus macaques. *Prog. Brain Res.* **64**, 117–131 (1986).
9. Kurata, K. Premotor cortex of monkeys: set- and movement-related activity reflecting amplitude and direction of wrist movement. *J. Neurophysiol.* **69**, 187–120 (1993).
10. Tanji, J., Taniguchi, K. & Soga, T. Supplementary motor area: neuronal response to motor instructions. *J. Neurophysiol.* **43**, 60–68 (1980).
11. Tanji, J. & Kurata, K. Contrasting neuronal activity in supplementary and precentral motor cortex of monkeys. I. Responses to instructions determining motor responses to forthcoming signals of different modalities. *J. Neurophysiol.* **53**, 129–141 (1985).
12. Alexander, G. E. Selective neuronal discharge in monkey putamen reflects intended direction of planned limb movements. *Exp. Brain Res.* **67**, 623–634 (1987).
13. Jaeger, D., Gilman, S. & Aldridge, J. W. Primate basal ganglia activity in a precued reaching task: preparation for movement. *Exp. Brain Res.* **95**, 51–64 (1993).
14. Dum, R. P. & Strick, P. L. The origin of corticospinal projection from premotor areas in the frontal lobe. *J. Neurosci.* **11**, 667–689 (1991).

15. Requin, J., Bonnet, M. & Semjen, A. In *Attention and Performance VI* (ed. Dornic, S.) 139–147 (Lawrence Erlbaum, Hillsdale, New Jersey, 1977).
16. Brunia, C. H., Scheirs, J. G. & Haagh, S. A. Changes of Achilles tendon reflex amplitudes during a fixed foreperiod of four seconds. *Psychophysiology* **19**, 63–70 (1982).
17. Komiya, T. & Tanaka, R. The differences in human spinal motoneuron excitability during the foreperiod of a motor task. *Exp. Brain Res.* **79**, 357–364 (1990).
18. Bonnet, M., Requin, J. & Stelmach, G. E. Changes in electromyographic responses to muscle stretch, related to the programming of movement parameters. *Electroencephalogr. Clin. Neurophysiol.* **81**, 135–151 (1991).
19. Burke, D., McKeon, B., Skuse, N. F. & Westerman, R. A. Anticipation and fusimotor activity in preparation for a voluntary contraction. *J. Physiol.* **306**, 337–348 (1980).
20. Gerilovsky, L., Struppler, A., Altman, H. & Velho, F. Spindle activity and monosynaptic reflex excitability during foreperiod. *Electroencephalogr. Clin. Neurophysiol.* **56**, 487–493 (1983).
21. Kuypers, H. G. & Brinkman, J. Precentral projections to different parts of the spinal intermediate zone in the rhesus monkey. *Brain Res.* **24**, 29–48 (1970).
22. Murray, E. A. & Coulter, J. D. Organization of corticospinal neurons in the monkey. *J. Comp. Neurol.* **195**, 339–365 (1981).
23. Toyoshima, K. & Sakai, H. Exact cortical extent of the origin of the corticospinal tract (CST) and the quantitative contribution to the CST in different cytoarchitectonic areas. A study with horseradish peroxidase in the monkey. *J. Hirnforsch.* **23**, 257–269 (1982).
24. He, S. Q., Dum, R. P. & Strick, P. L. Topographic organization of corticospinal projections from the frontal lobe: motor areas on the lateral surface of the hemisphere. *J. Neurosci.* **13**, 952–980 (1993).
25. Perlmuter, S. I., Maier, M. A. & Fetz, E. E. Activity of spinal interneurons and their effects on forearm muscles during voluntary wrist movements in the monkey. *J. Neurophysiol.* **80**, 2475–2494 (1998).
26. Kurata, K. & Wise, S. P. Premotor cortex of rhesus monkeys: set-related activity during two conditional motor tasks. *Exp. Brain Res.* **69**, 327–343 (1988).
27. Moll, L. & Kuypers, H. G. Premotor cortical ablations in monkeys: contralateral changes in visually guided reaching behavior. *Science* **198**, 317–319 (1977).
28. Sawaguchi, T., Yamane, I. & Kubota, K. Application of the GABA antagonist bicuculline to the premotor cortex reduces the ability to withhold reaching movements by well-trained monkeys in visually guided reaching task. *J. Neurophysiol.* **75**, 2150–2156 (1996).
29. Baumgartner, C. et al. Epileptic negative myoclonus: An EEG-single-photon emission CT study indicating involvement of premotor cortex. *Neurology* **46**, 753–758 (1996).

Acknowledgements

We thank S. Perlmuter for his help during the experiments; S. Wise for critical review of the manuscript; J. Garlid and L. Shupe for technical assistance; and K. Elias for editorial help.

Correspondence and requests for materials should be addressed to Y.P. (e-mail: yifat@u.washington.edu).

Polyamine-dependent facilitation of postsynaptic AMPA receptors counteracts paired-pulse depression

Andrei Rozov & Nail Burnashev

Max-Planck-Institut für medizinische Forschung, Abteilung Zellphysiologie/
Molekulare Neurobiologie, Jahnstraße 29, D-69120 Heidelberg, Germany

At many glutamatergic synapses in the brain, calcium-permeable α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate receptor (AMPA) channels mediate fast excitatory transmission^{1–6}. These channels are blocked by endogenous intracellular polyamines^{7–9}, which are found in virtually every type of cell^{10,11}. In excised patches, use-dependent relief of polyamine block enhances glutamate-evoked currents through recombinant and native calcium-permeable, polyamine-sensitive AMPAR channels¹². The contribution of polyamine unblock to synaptic currents during high-frequency stimulation may be to facilitate currents and maintain current amplitudes in the face of a slow recovery from desensitization or presynaptic depression^{12,13}. Here we show, on pairs and triples of synaptically connected neurons in slices, that this mechanism contributes to short-term plasticity in local circuits formed by presynaptic pyramidal neurons and postsynaptic multipolar interneurons in layer 2/3 of rat neocortex. Activity-dependent relief from polyamine block of postsynaptic calcium-permeable AMPARs in the interneurons either reduces

the rate of paired-pulse depression in a frequency-dependent manner or, at a given stimulation frequency, induces facilitation of a synaptic response that would otherwise depress. This mechanism for the enhancement of synaptic gain appears to be entirely postsynaptic.

To determine the contribution of polyamine (PA)-dependent facilitation to short-term plasticity of excitatory synaptic transmission, we selected synapses of pyramidal cells on multipolar interneurons in layer 2/3 of rat neocortex¹⁴. This excitatory connection is characterized by a high release probability¹⁴, and somatic AMPAR channels in multipolar interneurons are Ca^{2+} -permeable (unpublished observations). Hence, this connection is a probable site for PA-mediated effects.

We first tested whether AMPAR channels that are expressed in the soma of multipolar interneurons undergo PA-dependent facilitation. When two brief (1 ms) glutamate (1 mM) pulses at 40 Hz were applied to outside-out patches pulled from the soma of multipolar interneurons with standard PA-containing intracellular solution (see Methods), the ratio of the second current amplitude to the first one (I_2/I_1) was 0.6 ± 0.07 (mean $\pm$ s.d., $n = 3$). The current–voltage (I – V) relation for the glutamate-activated currents was doubly rectifying. With PA-free intracellular solution (see Methods), the ratio was significantly smaller (0.37 ± 0.05 , $n = 3$), and the I – V relation was linear (Fig. 1a). In contrast, in patches pulled from layer 2/3 pyramidal cells, neither the shape of I – V relation nor the degree of current reduction under the same experimental conditions was affected by the presence ($I_2/I_1 = 0.81 \pm 0.04$, $n = 3$) or absence ($I_2/I_1 = 0.79 \pm 0.06$, $n = 3$) of spermine in the recording pipette (Fig. 1b). Thus, AMPAR channels in the soma of multipolar interneurons, but not in pyramidal neurons, are sensitive to intracellular PAs and undergo facilitation similar to that reported for Ca^{2+} -permeable recombinant and native AMPAR channels¹².

We then tested whether synaptic AMPAR channels in multipolar interneurons were PA-sensitive. Simultaneous whole-cell recordings were made from pairs of synaptically connected layer 2/3 pyramidal and multipolar interneurons. Excitatory postsynaptic potentials (EPSPs), evoked by action potentials in presynaptic

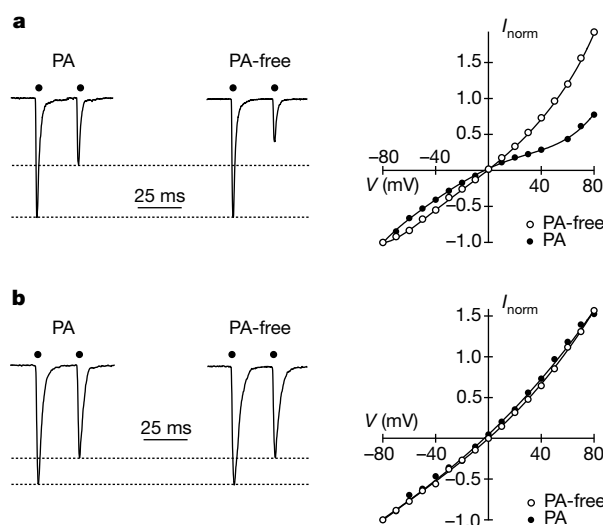


Figure 1 Polyamine-dependent facilitation counteracts AMPAR-mediated current desensitization in outside-out patches. **a**, Glutamate-activated current traces recorded from outside-out patches excised from multipolar interneurons with standard PA-containing and PA-free intracellular solutions. Membrane potential, -60 mV. Currents were normalized to the amplitude of the first response. Right, corresponding I – V relations for the first currents. For comparison, current amplitudes were normalized to the values at -80 mV. **b**, As in **a** except that outside-out patches were excised from pyramidal neurons.

pyramidal cells were measured in multipolar interneurons in control extracellular solution and with 50 μ M extracellular spermine. The amplitude of EPSPs measured in spermine was reduced to $44 \pm 1\%$ of that of the control ($n = 3$; Fig. 2a), indicating that synaptic AMPAR channels in multipolar interneurons are also PA-sensitive. In these experiments, PA-free intracellular solution was used. Similar experiments performed on pairs of synaptically connected pyramidal cells did not reveal a strong effect of extracellular spermine on the amplitude of the unitary EPSP. The EPSP amplitude was reduced to $87 \pm 3\%$ of the control value ($n = 3$; Fig. 2b), which is comparable to that reported for pyramidal cells in the amygdala⁶. We used this type of connection in this and further experiments as a control system to exclude nonspecific effects of PAs, and intracellular solutions used on synaptic transmission.

Simultaneous whole-cell recordings have shown that during repetitive (10 Hz) stimulation of presynaptic pyramidal cells in standard extracellular solution (2 mM extracellular Ca^{2+}) at both physiological and room temperature, EPSPs evoked in multipolar interneurons exhibit prominent paired-pulse depression (PPD)¹⁴. As synaptic AMPARs in this type of interneuron are PA-sensitive, one might expect that intracellular PAs may reduce the degree of depression in a similar fashion to that for currents in outside-out patches.

We measured PPD of excitatory postsynaptic currents (EPSCs) in multipolar interneurons in response to pairs of action potentials (40 Hz) elicited in pyramidal neurons at room temperature. Under these conditions, PPD, measured as a ratio of the averaged amplitudes of the second and first EPSCs (EPSC2/EPSC1) and recorded with PA-containing standard intracellular solution, was 0.6 ± 0.13 . After repatching the same multipolar interneurons with a pipette containing PA-free intracellular solution, the EPSC2/EPSC1 ratio decreased to 0.39 ± 0.13 ($n = 6$). At the same time, the amplitude of the first EPSC increased on average to $157 \pm 21\%$ of its original value ($n = 6$), which is consistent with the reduced block of AMPAR channels caused by washout of PAs from postsynaptic sites (Fig. 2c). Similar experiments performed on pairs of synaptically connected pyramidal cells did not reveal any effect of spermine on either the

amplitude of the first EPSC or on the synaptic depression (Fig. 2d). The amplitude of the first EPSC in PA-free solution was $95 \pm 8\%$, ($n = 6$) of that in PA-containing solution. The EPSC2/EPSC1 ratios at the interpulse interval of 25 ms were 0.48 ± 0.08 and 0.48 ± 0.09 ($n = 4$) with PA-free and PA-containing intracellular solutions, respectively. This is in line with insensitivity to intracellular PA block of the AMPAR channels expressed in pyramidal cells. As all manipulations with intracellular solutions that differ only in spermine concentration were made on postsynaptic cells, the effect of spermine on synaptic efficacy in multipolar interneurons appears to be entirely postsynaptic.

The effect of spermine on PPD in multipolar interneurons was dependent on the frequency of synaptic stimulation. At 40 Hz stimulation, PPD was reduced in the presence of spermine on average by 0.21 (a difference between PPD with PA-containing and PA-free solutions). At 10 Hz stimulation, the reduction in PPD was much less (0.09, Fig. 2c). In synaptically connected pyramidal cells, again there was no effect of spermine on EPSCs at any tested stimulation frequency (Fig. 2d).

To determine the mechanism underlying the relative enhancement of the second EPSC in the presence of PA in the recording pipette, we compared I - V relations for the first and the second EPSCs measured with the standard PA-containing intracellular solution using a paired-pulse protocol with 25 ms interpulse interval. Figure 3a shows I - V relations for EPSCs normalized to the EPSC amplitude at -80 mV and the original traces. The I - V relation for the first EPSC was sigmoidal, as expected for the PA-sensitive AMPAR channels mediating synaptic transmission; however, the I - V relation for the second EPSCs at negative potentials deviated from that for the first EPSCs, consistent with activity-dependent relief from PA block at negative potentials¹². Thus, partial relief from PA block during the first EPSC underlies the enhanced relative amplitude of the second one. In contrast, the I - V relations for the normalized first and the second EPSCs measured using the same experimental protocol but with PA-free intracellular solution were both almost linear and not significantly different from each other (Fig. 3b). Thus, in intact synapses, activity-dependent

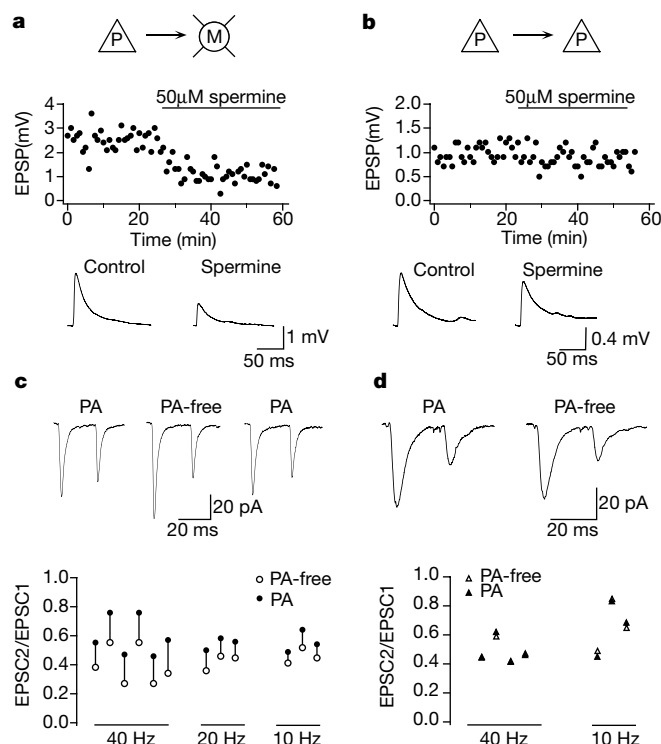


Figure 2 PAs reduce amplitude and PPD of synaptic responses in multipolar interneurons but not in pyramidal cells. **a, b**, Extracellular spermine (50 μ M) reduces amplitude of EPSPs evoked in multipolar interneurons (**a**) but not in pyramidal cells (**b**) during stimulation (0.1 Hz) of presynaptic pyramidal neurons. Each point on the scatter plot is the average of five consecutive sweeps. Traces represent the average of 100 EPSPs in control and in spermine at the steady-state level. **c, d**, Intracellular spermine (50 μ M in solution **a**, see Methods) reduces both the amplitude of the first EPSC and PPD evoked in multipolar interneurons but not in pyramidal cells during paired-pulse stimulation (40 Hz) of presynaptic pyramidal neurons. **c**, Upper traces, averaged EPSCs recorded from the same multipolar interneuron with standard PA-containing intracellular solution (left), after washout of spermine (middle), and after washing in 50 μ M spermine (right; see Methods for details). Plot shows pairwise comparison of the ratio of EPSC2/EPSC1 evoked by paired-pulse protocol recorded with the standard PA-containing and PA-free intracellular solutions in multipolar interneurons at different stimulation frequencies. **d**, As in **c** except that pyramidal neurons were used as target cells. The wash in of spermine is not shown.

relief from PA block of postsynaptic AMPAR channels underlies the reduction in PPD.

To quantify the frequency dependence of PA unblock, we measured recovery from depression of EPSPs in multipolar interneurons evoked by two action potentials elicited in pyramidal cells delayed by different time intervals. Figure 4a shows a clear difference in the time course of recovery from depression depending on the presence or absence of spermine in the postsynaptic recording pipette. After a 25 ms interval, almost 40% of depression was abolished by PA-dependent facilitation. The effect reduced gradually with longer time intervals and was not detected at 0.5 s. Complete recovery time in both cases was 5 s (not shown). Converting the time dependence of the ratio of EPSP amplitudes (Fig. 4a) to the frequency dependence (Fig. 4b) indicates that PA-dependent facilitation shifts the cut-off frequency of the synaptic signal filtering induced by presynaptic depression and postsynaptic AMPAR desensitization to higher frequencies, thus enhancing the dynamic range of synaptic transmission. By subtracting the two graphs one can estimate the real facilitating effect of spermine unblock and its frequency dependence. The effect of spermine on EPSPs was more pronounced at higher frequencies of delivery of action potentials (Fig. 4b). Thus, synaptic gain in multipolar interneurons might be dynamically regulated by postsynaptic AMPAR channels, depending on the firing rate of presynaptic pyramidal cells.

The experiments described so far were made at room temperature with 2 mM Ca^{2+} in the extracellular solution. Further experiments showed that under more physiological conditions, with an extracellular Ca^{2+} concentration of 1.6 mM¹⁵ and at 34 °C, PA-dependent facilitation can result in synaptic facilitation in synapses that would otherwise depress. EPSP2/EPSP1 amplitude ratios measured in multipolar interneurons with the standard PA-containing intracellular solution in the recording pipette ranged from 0.92 to 1.47, exhibiting, on average, pronounced paired-pulse facilitation (EPSP2/EPSP1 = 1.21 ± 0.15 , $n = 19$). With PA-free intracellular solution, PPD was observed (EPSP2/EPSP1 = 0.78 ± 0.09 , $n = 17$). To confirm that this difference is determined entirely by postsynaptic mechanisms, EPSPs were recorded simultaneously in two multipolar interneurons receiving synaptic inputs from the same pyramidal neurons. One of the multipolar interneurons was loaded with the standard PA-containing intracellular solution, the other with PA-free intracellular solution. Two action potentials (at 20 Hz) elicited in the pyramidal neuron evoked EPSPs in both the multipolar interneurons, however, the EPSP2/EPSP1 ratio of the two multipolar cells differed markedly. EPSPs recorded from the multipolar interneuron loaded with the PA-containing solution showed paired-pulse facilitation (PPF) or slight depression. In contrast, EPSPs recorded simultaneously from the multipolar interneuron loaded with PA-free intracellular solution always depressed (Fig. 4c). Various factors could contribute to the differences seen between the experiments performed at room temperature and those under more physiological conditions. First, lower extracellular Ca^{2+} concentration compared with that in the standard extracellular solution would be expected to reduce release probability¹⁶. Second, the higher temperature accelerates regeneration of the released vesicles¹⁷. Both these factors would reduce presynaptic depression by reducing vesicle depletion so that PA-dependent facilitation of postsynaptic AMPAR-mediated currents becomes crucial in determining the pattern of synaptic activity in multipolar interneurons. Thus, at certain stimulation frequencies, PA-dependent facilitation underlies synaptic facilitation.

To access the contribution of the endogenous PAs to short-term plasticity, we performed experiments in which the recording of EPSPs was started immediately after formation of the whole-cell mode in interneurons. For this, the search of the connected cell pairs was reversed so that first the projecting cell and then the target cell were patched. In these experiments, intracellular solution 'b' was used to slow down washout of endogenous PAs. As washout of

endogenous PAs from postsynaptic sites is not instant, at the initial phase of the recording one can expect concentration of endogenous PAs to be close to normal. We monitored changes in amplitude of the first EPSP and paired-pulse ratio (PPR) throughout 15-min recordings. The amplitude of the first EPSP gradually increased (Fig. 4d, closed circles) consistent with the washout of endogenous PAs. At the same time, the EPSP2/EPSP1 ratio estimated from averaged EPSPs measured over 0–7 min and 8–15 min significantly decreased from 1.39 ± 0.27 to 0.84 ± 0.08 ($n = 5$; paired t -test, $P = 0.01$), indicating that after partial washout of endogenous PAs the PPF changed to PPD. Similar experiments with intracellular

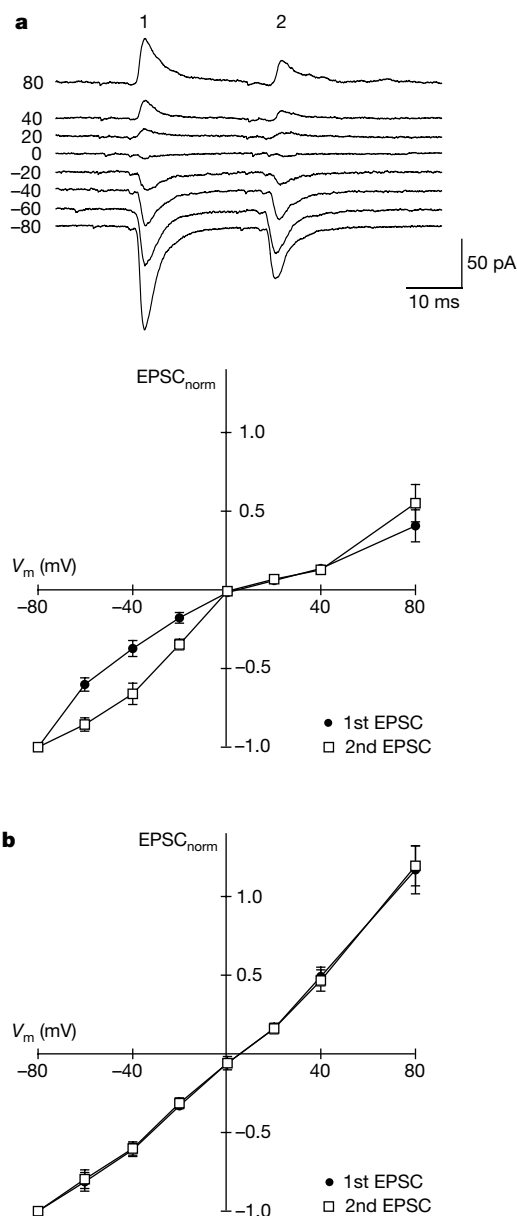


Figure 3 Relief from PA block underlies enhancement of the second EPSC. **a**, Upper traces, EPSCs evoked by paired-pulse protocol (25 ms interpulse interval) recorded with the standard PA-containing intracellular solution in multipolar interneurons at membrane potentials shown (in mV) on the left. Lower plot, I – V relations for the first and second EPSCs. For comparison, current amplitudes were normalized to the values at -80 mV for each set of EPSCs. **b**, Normalized I – V relations for the first and second EPSCs recorded as in **a** but with PA-free intracellular solution in multipolar interneurons. Note that both I – V relations are almost identical. Each point in **a** and **b** represents average amplitudes recorded from 5–7 cell pairs. Error bars, s.e.m.

solution 'b' containing 50 μM spermine to compensate for loss of PAs showed the opposite trend in change of the amplitude of the first EPSP (Fig. 4d, open circles), and the PPR did not change significantly (1.24 ± 0.19 and 1.33 ± 0.23 , $n = 4$, $P = 0.29$). Notably, in 2 out of a total of 11 pairs, PPD was observed at the initial phase of the recordings. As with these cells the recording broke down within the first 5–7 min, EPSPs at later times could not be recorded. These cell pairs were not included in statistics. Together, these data indicate that endogenous PAs present in multipolar interneurons may produce similar effects on EPSP amplitude and PPR to those of exogenously added spermine. Thus, during prolonged whole-cell recordings, PAs can be washed out and their possible contribution to EPSP amplitude and PPD might be overlooked.

Polyamines are normal intracellular constituents found in almost any cell type^{10,11,18}. In neurons that express Ca^{2+} -permeable, PA-sensitive AMPAR channels, endogenous PAs reduce the amplitude of the AMPAR-mediated synaptic response (for example, Figs 2c, 4d). On the other hand, this initial block allows the enhancement of the synaptic response during the high-frequency stimulation that arises from activity-dependent relief from the PA block. The relative contribution of PA-dependent facilitation to short-term synaptic plasticity will depend on presynaptic depletion of vesicles and postsynaptic AMPAR desensitization. In local circuits in layer 2/3

of rat visual cortex formed by presynaptic pyramidal neurons and postsynaptic multipolar interneurons under more physiological conditions, the interplay between mechanisms contributing to short-term synaptic plasticity shifts in favour of PA-dependent facilitation so that this mechanisms accounts for synaptic facilitation. Thus, PA-dependent facilitation provides an entirely postsynaptic mechanism of dynamic regulation of synaptic gain that may determine target-cell-specific differences in synaptic transmission in neuronal circuits. In addition to this important functional consideration, the significant postsynaptic influence on PPRs means that previous assumptions concerning the presynaptic locus of such changes^{19–21} must be reconsidered for synapses where PA-sensitive AMPA receptors are expressed. □

Methods

Transverse neocortical slices of 300 μm thickness were prepared from the brains of 14-day-old Wistar rats killed by decapitation. During recordings, slices were maintained at room temperature ($22\text{--}24^\circ\text{C}$) in standard Ringer extracellular solution (Biometra) consisting of (in mM) 125 NaCl, 2.5 KCl, 25 H_2CO_3 , 1.25 NaH_2PO_4 , 2 CaCl_2 and 1 MgCl_2 , pH 7.2. Some experiments were performed at 34°C and with 1.6 mM CaCl_2 . Cells were identified visually using infrared differential contrast video microscopy²² and by their firing pattern following depolarizing current injection¹⁴. Whole-cell voltage and/or current recordings were performed simultaneously from two or three synaptically connected neurons using pipettes with resistance of 5–8 M Ω . Presynaptic pyramidal cells were stimulated with 10–40 Hz trains of two suprathreshold current pulses. Trains were

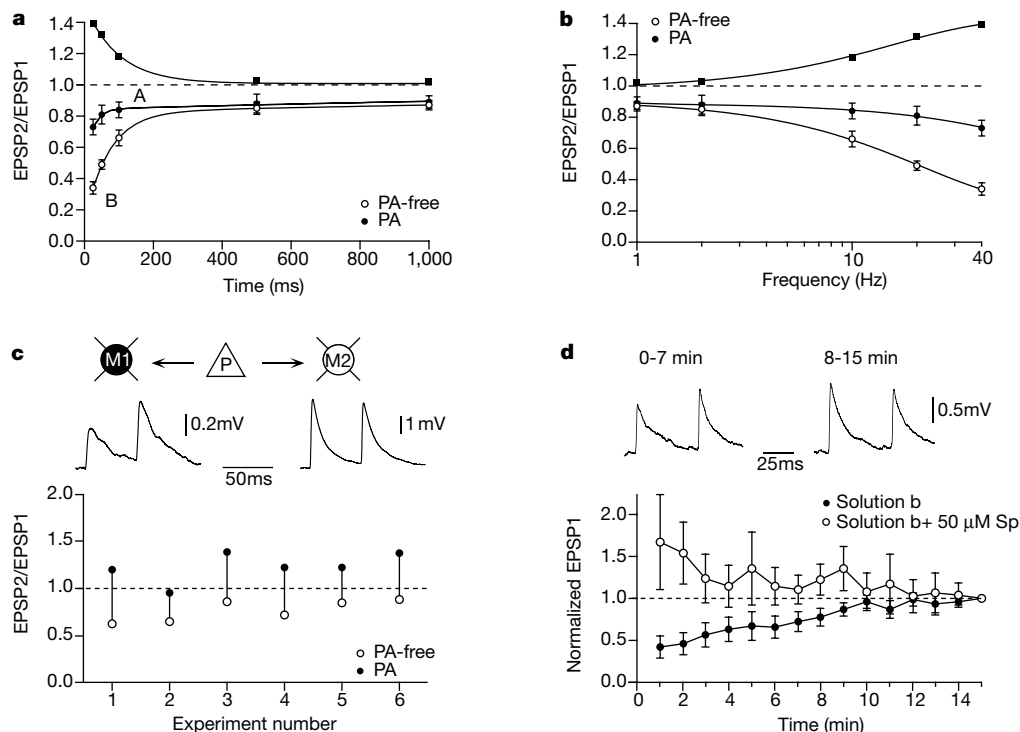


Figure 4 Physiological properties of PA-dependent facilitation of synaptic AMPAR channels in multipolar interneurons. **a**, **b**, Activity-dependent relief from PA block is frequency dependent. **a**, Recovery from PPD of EPSCs evoked by paired-pulse protocol with variable interpulse intervals recorded with the standard PA-containing (A) and PA-free (B) intracellular solutions in multipolar interneurons ($22\text{--}24^\circ\text{C}$ and 2 mM extracellular Ca^{2+}). Filled squares, obtained by the equation $A - B + 1$, represent recovery from PA-dependent facilitation. **b**, The same data as in **a** plotted against frequency in a semilogarithmic scale. PA-induced facilitation maintains synaptic gain at higher frequencies. **c**, PA-dependent facilitation is entirely postsynaptic. Simultaneous whole-cell recordings were made at 34°C and in 1.6 mM extracellular Ca^{2+} from a triplet (schematically shown on top) in which the same pyramidal neuron (P) innervated two multipolar interneurons (M1, M2). M1 was loaded with the standard PA-containing, M2 with the PA-free intracellular solutions. EPSPs evoked by two action potentials elicited in

the pyramidal cell at 20 Hz facilitated in M1 (left trace) and depressed in M2 (right trace). Plot shows pairwise comparison of EPSP2/EPSP1 ratios measured from 6 triplets. Connected symbols represent amplitude ratios of EPSPs evoked simultaneously in multipolar interneurons loaded with the standard PA-containing and with the PA-free intracellular solutions following 20 Hz stimulation of the same presynaptic pyramidal neuron. **d**, Endogenous PAs induce synaptic facilitation. Averaged EPSPs recorded from the same multipolar interneuron in paired-pulse protocol (20 Hz) during first 7 min after whole-cell formation (left trace) and within 8–15 min (right trace). Intracellular solution b was used in the recording pipette (34°C and 1.6 mM extracellular Ca^{2+}). Action-potential pairs were delivered every 15 s. Plot shows normalized amplitudes of the first EPSPs (each point represents the average of 4 consecutive sweeps) recorded during 15 min following formation of the whole-cell mode with intracellular solution b without (5 pairs) or with 50 μM spermine (4 pairs). Error bars, s.e.m.

delivered at intervals of 12–30 s. Voltage and current traces shown are averages of 50–100 sweeps. Recordings were made using an EPC-7 amplifier (List Elektronik). Stimulus delivery and data acquisition were performed using PULSE software (HEKA Elektronik). All analyses were performed using IGOR PRO software (Wavemetrics). Outside-out patches were pulled from the target cells and brief pulses (1 ms) of glutamate (1 mM) were applied using a Piezo-controlled (P 245.70, Physik Instrumente) fast application system with a double-barrel application pipette²³. AMPAR-mediated currents in patches were recorded in the presence of 100 μ M D-AP-5. Averaged data are given as mean $\pm$ s.d. unless otherwise noted.

The two intracellular solutions comprised (in mM): (a) 105 K-gluconate, 20 K₂-ATP, 10 KCl, 10 phosphocreatine, 0.3 GTP, 4 Mg-ATP, 10 HEPES-KOH, pH 7.2; or (b) 105 K-gluconate, 30 KCl, 4 Mg-ATP, 10 phosphocreatine, 0.3 GTP and 10 HEPES-KOH, pH 7.3. Solution a was used to achieve PA-free conditions; spermine (50 μ M) was added to a to give a free concentration of roughly 3.3 μ M¹⁸. Solution a with 50 μ M spermine was used as a standard PA-containing intracellular solution to prevent uncontrollable change in intracellular PA concentration during prolonged recordings. For measurements of synaptic *I*-*V* relations, K-gluconate and KCl in solution a were replaced by Cs-gluconate and CsCl, respectively. Typically recordings started 10–15 min after making the whole-cell configuration to allow complete loading of the terminals (but see Fig. 4d). In experiments shown in Fig. 2c, after control recordings with the standard PA-containing intracellular solution, the target cells were repatched with the pipette containing solution b for 15–20 min to accelerate washout of bound spermine by increasing ATP outflow gradient. After 30 min, the same cell was repatched with the pipette containing PA-free solution a. In projecting pyramidal cells in all experiments intracellular solution b was used.

Received 11 May; accepted 2 August 1999.

- Gu, J. G., Albuquerque, C., Lee, C. J. & MacDermott, A. B. Synaptic strengthening through activation of Ca²⁺-permeable AMPA receptors. *Nature* **381**, 793–796 (1996).
- McBain, C. J. & Dingledine, R. Heterogeneity of synaptic glutamate receptors on CA3 stratum radiatum interneurons of rat hippocampus. *J. Physiol. (Lond.)* **462**, 373–392 (1993).
- Otis, T. S., Raman, I. M. & Trussell, L. O. AMPA receptors with high Ca²⁺ permeability mediate synaptic transmission in the avian auditory pathway. *J. Physiol. (Lond.)* **482**, 309–315 (1995).
- Isa, T., Itazawa, S., Iino, M., Tsuzuki, K. & Ozawa, S. Distribution of neurones expressing inwardly rectifying and Ca²⁺-permeable AMPA receptors in rat hippocampal slices. *J. Physiol. (Lond.)* **491**, 719–733 (1996).
- Barnes-Davies, M. & Forsythe, I. D. AMPA receptor-mediated synaptic currents rectify with internal spermine in rat MNTB neurones *in vitro*. *J. Physiol. (Lond.)* **495**, 44P (1996).
- Mahanty, N. K. & Sah, P. Calcium-permeable AMPA receptors mediate long-term potentiation in interneurons in the amygdala. *Nature* **394**, 683–687 (1998).
- Koh, D.-S., Burnashev, N. & Jonas, P. Block of native Ca²⁺-permeable AMPA receptors in rat brain by intracellular polyamines generates double rectification. *J. Physiol. (Lond.)* **486**, 305–312 (1995).
- Kamboj, S. K., Swanson, G. T. & Cull-Candy, S. G. Intracellular spermine confers rectification on rat calcium-permeable AMPA and kainate receptors. *J. Physiol. (Lond.)* **486**, 297–303 (1995).
- Bowie, D. & Mayer, M. L. Inward rectification of both AMPA and kainate subtype glutamate receptors generated by polyamine-mediated ion channel block. *Neuron* **15**, 453–462 (1995).
- Shaw, G. G. The polyamines in the central nervous system. *Biochem. Pharmacol.* **28**, 1–6 (1979).
- Pegg, A. E. Recent advances in the biochemistry of polyamines in eukaryotes. *Biochem. J.* **234**, 249–262 (1986).
- Rozov, A., Zilberter, Yu., Wollmuth, L. P. & Burnashev, N. Facilitation of currents through rat Ca²⁺-permeable AMPA receptor channels by activity-dependent relief from polyamine block. *J. Physiol. (Lond.)* **511**, 361–377 (1998).
- McBain, C. J. A short-term mechanism of plasticity for interneurons? *J. Physiol. (Lond.)* **511**, 331 (1998).
- Reyes, A. et al. Target-cell-specific facilitation and depression in neocortical circuits. *Nature Neurosci.* **1**, 279–285 (1998).
- Jones, H. C. & Keep, R. F. Brain fluid calcium concentration and response to acute hypercalcaemia during development in the rat. *J. Physiol. (Lond.)* **402**, 579–593 (1988).
- Thomson, A. M. Activity-dependent properties of synaptic transmission at two classes of connections made by rat neocortical pyramidal axons *in vitro*. *J. Physiol. (Lond.)* **502**, 131–147 (1997).
- Stevens, C. F. & Wang, Y. Facilitation and depression at single central synapses. *Neuron* **14**, 795–802 (1995).
- Watanabe, S., Kusama-Eguchi, K., Kobayashi, H. & Igarashi, K. Estimation of polyamine binding to macromolecules and ATP in bovine lymphocytes and rat liver. *J. Biol. Chem.* **266**, 20803–20809 (1991).
- Zucker, R. S. Short-term synaptic plasticity. *Annu. Rev. Neurosci.* **12**, 13–31 (1989).
- Zucker, R. S. Calcium and short-term synaptic plasticity. *Netherlands J. Zool.* **44**, 495–512 (1994).
- Fisher, S. A., Fischer, T. M. & Carew, T. J. Multiple overlapping processes underlying short-term synaptic enhancement. *Trends Neurosci.* **20**, 170–177 (1997).
- Stuart, G. J., Dodt, H.-U. & Sakmann, B. Patch-clamp recordings from the soma and dendrite of neurons in brain slices using infrared video microscopy. *Pflügers Arch.* **423**, 511–518 (1993).
- Colquhoun, D., Jonas, P. & Sakmann, B. Action of brief pulses of glutamate on AMPA/kainate receptors in patches from different neurones of rat hippocampal slices. *J. Physiol. (Lond.)* **458**, 261–287 (1992).

Acknowledgements

We thank B. Sakmann and P. H. Seeburg for support, and F. A. Edwards and S. Robertson for reading and comments on the manuscript.

Correspondence and requests for materials should be addressed to N.B. (e-mail: nail@sunny.mpimf-heidelberg.mpg.de).

Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds

Aimée Zúñiga*†‡, Anna-Pavlina G. Haramis*‡§, Andrew P. McMahon|| & Rolf Zeller†‡

† Department of Developmental Biology, Faculty of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

‡ EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany

|| Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

* These authors contributed equally to this work.

Outgrowth and patterning of the vertebrate limb are controlled by reciprocal interactions between the posterior mesenchyme (polarizing region) and a specialized ectodermal structure, the apical ectodermal ridge (AER)¹. Sonic hedgehog (SHH) signalling by the polarizing region modulates fibroblast growth factor (FGF)4 signalling by the posterior AER, which in turn maintains the polarizing region (SHH/FGF4 feedback loop)^{2,3}. Here we report that the secreted bone-morphogenetic-protein (BMP) antagonist *Gremlin*⁴ relays the SHH signal from the polarizing region to the AER. Mesenchymal *Gremlin* expression is lost in limb buds of mouse embryos homozygous for the *limb deformity* (*ld*) mutation, which disrupts establishment of the SHH/FGF4 feedback loop^{5–7}. Grafting *Gremlin*-expressing cells into *ld* mutant limb buds rescues *Fgf4* expression and restores the SHH/FGF4 feedback loop. Analysis of *Shh*-null mutant embryos^{8,9} reveals that SHH signalling is required for maintenance of *Gremlin* and *Formin* (the gene disrupted by the *ld* mutations)^{10,11}. In contrast, *Formin*, *Gremlin* and *Fgf4* activation are independent of SHH signalling. This study uncovers the cascade by which the SHH signal is relayed from the posterior mesenchyme to the AER and establishes that *Formin*-dependent activation of the BMP antagonist *Gremlin* is sufficient to induce *Fgf4* and establish the SHH/FGF4 feedback loop.

In *ld* mutant limb buds, *Shh* is activated normally in the posterior mesenchyme, but subsequent upregulation and distal propagation are disrupted owing to a mesenchymal defect which blocks *Fgf4* activation in the *ld* mutant AER^{6,7,12}. This failure to establish the SHH/FGF4 feedback loop results in severe patterning defects which primarily affect distal limb skeletal elements (digits). These results indicated either that SHH signalling by the polarizing region remains below the threshold necessary to induce *Fgf4* in the AER⁶ or that *Formin* is involved in relaying the SHH signal from the limb-bud mesenchyme to the AER⁷. To discriminate between these two possibilities, aggregates of SHH-expressing cells¹³ were grafted into wild-type (Fig. 1b) and *ld* mutant mouse forelimb buds (Fig. 1c) and *Fgf4* expression was analysed after culturing embryonic trunks for 19–22 h (see Methods). As expected from previous studies in chick limb buds^{2,3}, anterior grafts of SHH-expressing cells induce an ectopic *Fgf4* domain in wild-type mouse limb buds (compare Fig. 1a and 1b). In contrast, the same SHH-expressing cells fail to induce *Fgf4* in the AER of *ld* mutant limb buds (Fig. 1c), indicating a defect in transducing the SHH signal. To determine whether *ld* mutant limb bud mesenchymal cells are at all responsive to SHH signalling, we examined the transcriptional activation of two known mesenchymal targets, the transcription factor *Gli*¹⁴ and the SHH receptor *Ptc*¹⁵ (Fig. 1d–f). Ectopic SHH signalling induces *Gli* to similar levels in both wild-type (Fig. 1d) and *ld* mutant limb

§ Present address: Department of Anatomy, University College London, Gower Street, London WC1E 6BT, U.K.

mesenchyme (Fig. 1e). Likewise, *Ptc* transcription is upregulated in response to ectopic SHH signalling in *ld* mutant limb buds (Fig. 1f). These results show that the initial mesenchymal response to SHH signalling is normal, whereas *Formin* is required to relay this signal to the AER.

During initiation of limb-bud development, *Formin* (*Fmn*) is predominantly expressed by the AER^{5,12,16} and at much lower, but significant, levels by mesenchymal cells^{6,17} (not detectable by *in situ* hybridization). Subsequently, *Formin* expression is upregulated in posterior-distal limb-bud mesenchyme (Fig. 1g) adjacent to the SHH-expressing polarizing region (Fig. 1h)¹⁶, indicating possible SHH-dependent transcriptional upregulation. Indeed, anterior grafts of SHH-expressing cells¹³ induce ectopic *Formin* expression in distal limb-bud mesenchyme (Fig. 1i) in a domain often match-

ing the extent of ectopic *Fgf4* in the AER (compare Fig. 1b and 1i). The results shown in Fig. 1 indicate that a secondary mesenchymal signal acting downstream of nuclear *Formin*¹⁶ is missing in *ld* mutant limb buds.

Analysis of SHH signalling during chick limb development indicates that BMP2 is an effector of SHH signalling^{1-3,18}. Consistent with low levels of SHH expression, reduced levels of *Bmp2* transcripts are detected in *ld* mutant limb buds, whereas *Bmp4* and *Bmp7* expression is not affected during early limb-bud development (data not shown). However, grafts of BMP2-expressing cells¹⁸ fail to restore *Fgf4* expression in *ld* mutant limb buds (data not shown). In contrast, inhibition of BMP signalling by secreted antagonists¹⁹ may positively regulate *Fgf4* expression and AER morphology during chick limb-bud development²⁰. In addition, the BMP antagonist *Noggin* functions downstream of SHH signalling during somite patterning and neural tube ventralization²¹.

Comparative analysis of several known BMP antagonists in wild-type and *ld* mutant limb buds (see Methods) resulted in the identification of *Gremlin*, a member of the DAN/cerberus gene family⁴, as a candidate for the missing signal (Fig. 2). The murine *Gremlin* gene is expressed by the posterior limb-bud mesenchyme (Fig. 2a-c) in a domain similar to that of *Formin* (compare Figs 1g and 2a). In contrast, *Gremlin* is not expressed in *ld* mutant limb buds (Fig. 2d-f and data not shown). As *Gremlin* expression in posterior mesenchyme borders the *Shh* domain (Fig. 2g), its expression might, like that of *Formin* (Fig. 1h, i), be regulated by SHH. Indeed, ectopic SHH signalling induces *Gremlin* transcription in wild-type (Fig. 2h) but not in *ld* mutant limb-bud mesenchyme (Fig. 2i). These results indicate that SHH-responsive limb-bud mesenchymal cells must activate *Formin* (Fig. 1i) to reach competence for *Gremlin* expression (Fig. 2h).

Simultaneous detection of *Fgf4* and *Gremlin* transcripts in wild-type limb buds (Fig. 2j-l) shows that the antero-posterior extent of the *Gremlin* domain closely matches that of *Fgf4* in the overlying AER (Fig. 2j-l), as expected for a positive regulatory signal. To assess directly the potential role of *Gremlin* as a downstream mediator of SHH signalling and as an inducer of *Fgf4*, *Gremlin* (GRE)-expressing cells (see Methods) were grafted into *ld* mutant limb buds (Fig. 3a-f). Indeed, both anterior and posterior grafts (Fig. 3a, b) but not mock-transfected cells (Fig. 3c) always induce prominent *Fgf4* expression in the *ld* mutant AER. Most interestingly, posterior grafts of GRE-expressing cells (Fig. 3e) also promote upregulation and maintenance of endogenous *Shh* expression (compare Fig. 3e with 3d; Fig. 3f), whereas anterior grafts do not induce *Shh* expression in *ld* (Fig. 3d) or wild-type limb buds *de novo* (data not shown). Furthermore, the late expression domains of 5'HoxD genes in limb buds are also regulated by SHH signalling¹ and are affected in *ld* mutant limb buds⁶. However, ectopic *Gremlin* expression in either wild-type or *ld* mutant limb buds neither induces nor alters 5'HoxD expression (data not shown). These results indicate that *Gremlin* functions primarily to relay the SHH signal from the posterior limb-bud mesenchyme to the AER. In *ld* mutant limb buds, the SHH/FGF4 feedback loop is restored, probably as a consequence of *Fgf4* expression's being triggered in response to *Gremlin* produced by cells grafted into posterior mesenchyme.

To confirm that extracellular BMP antagonism does result in *Fgf4* activation in *ld* mutant limb buds, we investigated the inducing potential of the BMP antagonist *Noggin*. *Noggin* has no sequence similarity to *Gremlin*⁴, its limb mesenchymal expression has no similarity to that of *Formin* (data not shown) and the limb phenotypes of *Noggin*-deficient mouse embryos argue against an essential role for *Noggin* in establishment of the SHH/FGF-4 feedback loop²². As for *Gremlin* (Fig. 3a-f), anterior grafts of *Noggin*-expressing cells²³ induce *Fgf4* but not *Shh* expression (Fig. 3g and data not shown; see also ref. 20), whereas posterior *Noggin* grafts also promote upregulation and maintenance of endogenous *Shh*

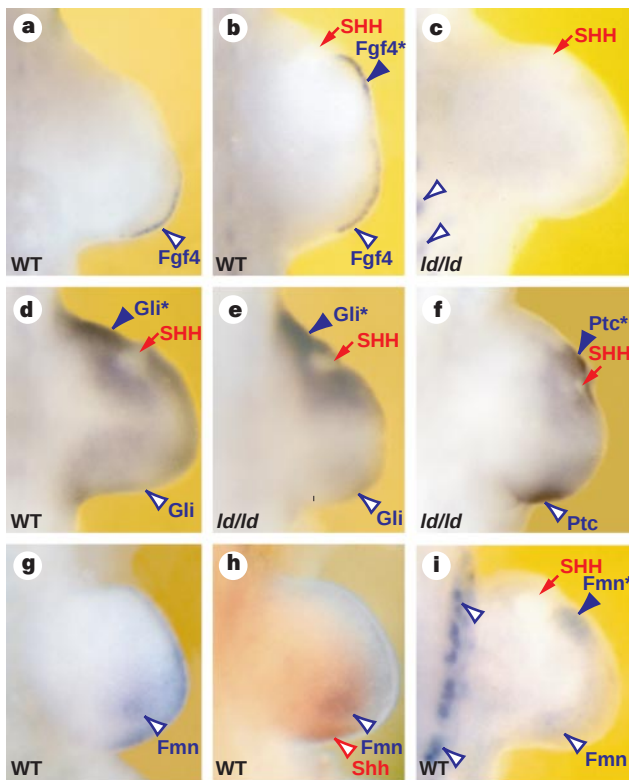


Figure 1 *Formin* (*Fmn*) is a target of SHH signalling in the limb-bud mesenchyme and is essential to relay the SHH signal from the mesenchyme to the AER. Red arrows, approximate positions of SHH-expressing cell aggregates after culturing; asterisks and filled arrowheads, ectopically induced expression domains; open arrowheads, endogenous domains. Posterior is to the bottom and distal to the right. Mouse forelimb buds (embryonic day (E) 10.25; 32–34 somites) received grafts of SHH-expressing cells¹³ and were cultured *in vitro* for 19–22 h (except **i**). **a**, Non-grafted wild-type forelimb bud. **b**, Wild-type forelimb bud (collateral to **a**). Ectopic *Fgf4* transcripts (*Fgf4**) are induced in the anterior AER by SHH-expressing cells. **c**, *ld* mutant (*ld/ld*) forelimb bud. SHH-expressing cells cannot induce *Fgf4* expression. Open arrowheads, endogenous *Fgf4* transcripts in somites. **d**, Wild-type forelimb bud. High levels of *Gli* transcripts (*Gli**) are induced in response to SHH-expressing cells. **e**, *ld* mutant forelimb bud. *Gli* transcripts similar to wild-type (**d**) are induced in response to SHH-expressing cells. **f**, *ld* mutant forelimb bud. *Ptc* transcription (*Ptc**) is also upregulated in response to SHH signalling. **g**, Mesenchymal *Fmn* mRNA distribution in a wild-type forelimb bud (E10.5; 35 somites). **h**, Parallel detection of *Fmn* (purple-red) and *Shh* transcripts (orange-red) in a wild-type forelimb bud (35 somites). **i**, Cultured wild-type forelimb bud. *Fmn* transcripts (*Fmn**) are induced in distal mesenchyme in response to SHH-expressing cells. The forelimb bud shown was cultured for 24 h to detect ectopic *Fmn* transcripts best. Weaker induction is detectable after 20 h (data not shown). Open arrowheads, endogenous *Fmn* transcripts in posterior limb mesenchyme and dorsal root ganglia.

expression (Fig. 3i). The importance of directly antagonizing BMP activity to permit *Fgf4* induction was further studied by placing BMP2-expressing cells¹⁸ close to a *Noggin* graft in *ld* mutant limb buds (Fig. 3h). Most *ld* mutant limb buds receiving such double grafts do not express *Fgf4* (Fig. 3h, compare with control shown in Fig. 3g). However, posterior grafts of BMP2-expressing cells into wild-type limb buds fail to suppress *Fgf4* expression (data not shown). Therefore, these results indicate that inhibition of BMP is essential to render the AER competent to activate *Fgf4* transcription. Together, the results shown in Figs 1–3 uncover the cascade by

which the SHH signal is relayed from the posterior limb-bud mesenchyme to the AER.

To establish definitively that mesenchymal *Formin* and *Gremlin* are dependent on SHH signalling, we analysed their expression in limb buds of *Shh* null mutant (*Shh*^{-/-}) mouse embryos^{8,9} (Fig. 4). Comparative analysis of *Shh* mutant and wild-type limb buds shows that early *Formin* expression is unaffected (Fig. 4a, b), whereas subsequent *Formin* upregulation in posterior-distal mesenchyme is completely blocked (Fig. 4g, h). These results are consistent with *Formin*'s functioning downstream of SHH during establishment of

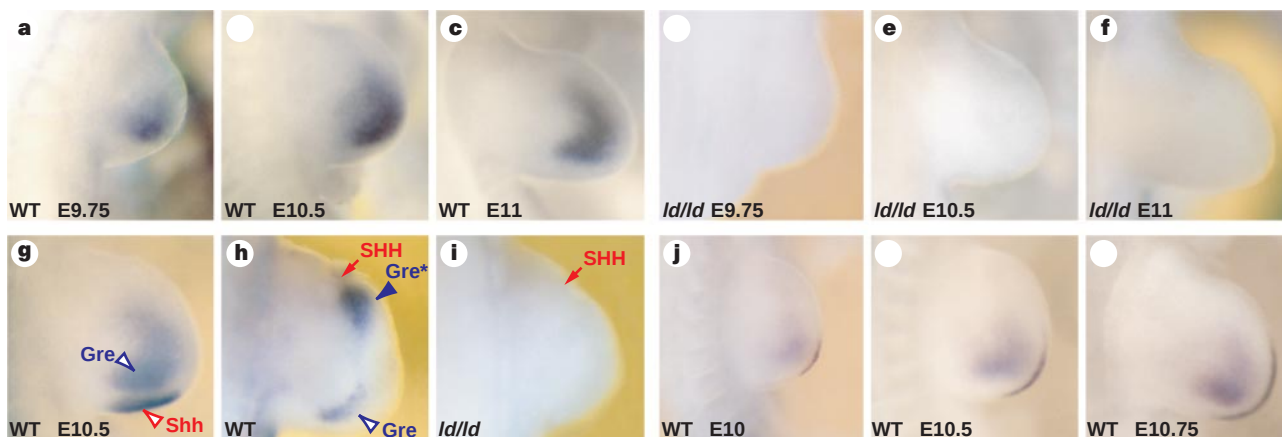


Figure 2 The BMP antagonist *Gremlin* (*Gre*) is a *Fmn*-dependent target of SHH signalling and is expressed by posterior-distal limb-bud mesenchymal cells. **a–f**, *Gre* transcript distribution in wild-type (**a–c**) and *ld* mutant forelimb buds (**d–f**) during E9.75 (28–29 somites; **a, d**), E10.5 (34–36 somites; **b, e**) and E11 (40 somites; **c, f**). *Gre* expression in wild-type limb buds is restricted to the mesenchyme (**a–c**), whereas no *Gre* expression is detected in the *ld* mutant limb buds (**d–f**). However, expression was normal in somite

derivatives (data not shown). **g**, Wild-type forelimb bud (E10.5; 35 somites). Simultaneous detection of *Gre* and *Shh* transcripts. **h, i**, SHH-expressing cells induce high levels of ectopic *Gre* transcription (*Gre*^{*}) in wild-type (**h**) but not *ld* mutant forelimb buds (**i**) after 19–22 h in culture. **j–l**, Simultaneous detection of *Gre* and *Fgf4* transcripts in wild-type forelimb buds during E10 (31 somites; **j**), E10.5 (34 somites; **k**) and E10.75 (37–38 somites, **l**).

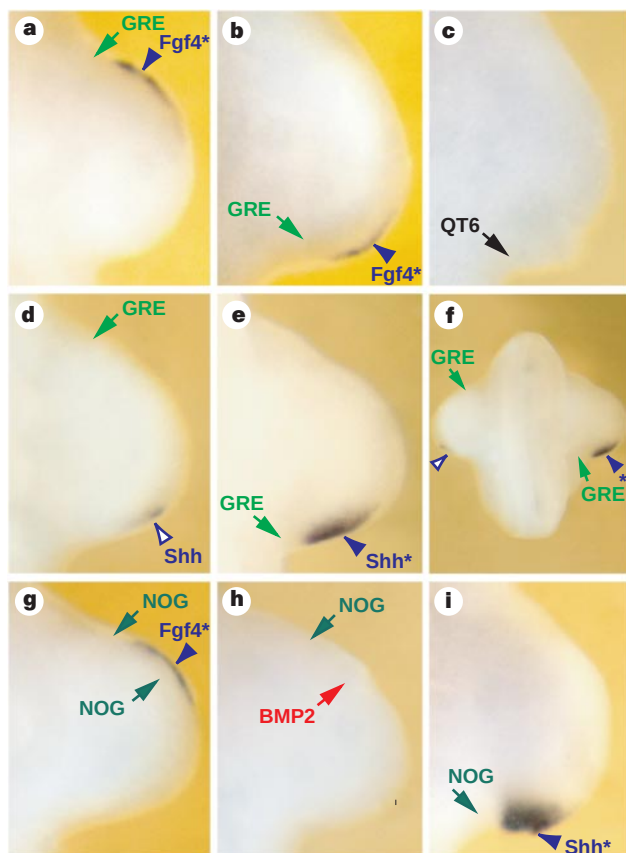


Figure 3 Inhibition of BMP activity by *Gremlin* restores the SHH/FGF-4 feedback loop in *ld* mutant limb buds. All grafted limb buds shown are forelimb buds of *ld* mutant embryos (32–34 somites) cultured for 19–22 h. Arrows, approximate positions of grafts after culturing; GRE (green), *Gremlin*-expressing cell aggregates; QT6 (black), mock-transfected parental QT6; NOG (dark green), *Noggin*-expressing cells²³; BMP2 (red), human-BMP2-expressing cells¹⁸. Filled arrowheads and asterisks mark induced transcripts; open arrowheads, endogenous expression domains. **a, b**, *Fgf4* is activated in the *ld* mutant AER in response to an anterior (**a**) or posterior graft (**b**) of GRE-expressing cells. **c**, *Fgf4* cannot be activated by mock-transfected QT6 cells. **d**, An anterior graft of GRE-expressing cells fails to induce *Shh* transcription *de novo* in *ld* mutant limb mesenchyme. Note low levels of endogenous *Shh* transcripts (open arrowhead) in posterior mesenchyme. **e**, Posterior graft of GRE-expressing cells induces upregulation of *Shh* transcripts. **f**, Low-magnification view of the *ld* mutant embryonic trunk whose forelimb buds are shown in **d** and **e**. **g**, Grafts of *Noggin* cells also trigger *Fgf4* expression. Two grafts of NOG-expressing cells were placed next to one another. This forelimb bud serves as a control for that shown in **h**. *Fgf4* expression is identical in *ld* mutant limb buds having received only one NOG graft (data not shown). **h**, Placement of a BMP2-expressing cell aggregate next to a NOG graft blocks *Fgf4* activation in the *ld* mutant AER. **i**, Posterior grafts of NOG-expressing cells also promote upregulation of *Shh* transcription comparable to GRE grafts (compare **i** with **e, f**).

the SHH/FGF4 feedback loop⁷. Early during limb-bud outgrowth, *Gremlin* transcription is induced in the posterior limb-bud mesenchyme (Fig. 4c). As for *Formin*, initial *Gremlin* expression is normal in *Shh* mutant limb buds (Fig. 4d). However, in contrast to wild-type limb buds (Figs 2a–c, 4i) *Gremlin* expression is rapidly lost from *Shh* mutant limb-bud mesenchyme (Fig. 4k). These results show that posterior signals other than SHH (compare Fig. 4d with Fig. 2h) can induce *Gremlin*, but that SHH signalling is essential to upregulate and maintain *Gremlin* expression in posterior-distal mesenchyme (compare Fig. 4k with Fig. 2j–k). In agreement with the SHH-independent activation of *Formin* and *Gremlin*, *Fgf4* is also induced in the AER of *Shh* mutant limb buds (Fig. 4f). Again, subsequent upregulation and distal-anterior expansion of the *Fgf4* domain are completely disrupted in *Shh*-deficient limb buds (compare Fig. 4m and 4l). The results shown in Figs 3 and 4 show that *Gremlin* is sufficient to induce *Fgf4* and indicate that mesenchymal *Gremlin* is probably necessary to both induce and propagate *Fgf4* expression during normal limb-bud development (Fig. 5).

The emerging cascade that triggers *Fgf4* expression in the posterior AER and leads to establishment of the SHH/FGF4 feedback loop is summarized in Fig. 5. An unknown posterior activity (PA), already present in the polarized limb field^{1,12}, acts independently and probably upstream of SHH to induce *Gremlin* in the limb-bud

mesenchyme (Fig. 5, upper panel). The complete loss of *Gremlin* expression in *ld* mutant limb buds (Fig. 3d–f; in contrast to *Shh*^{−/−} limb buds, Fig. 4d) indicates that *Gremlin* induction is *Formin*-dependent (Fig. 5, upper panel). Indeed, *Formin* is expressed at low levels by limb-bud mesenchymal cells^{16,17} before its prominent upregulation in posterior-distal mesenchyme (Figs 1g, 4g). In contrast to previous proposals¹, this study also establishes that *Fgf4* expression is induced in the AER independently of SHH signalling, probably as a consequence of early mesenchymal BMP antagonism mediated by *Gremlin* (Fig. 5, upper panel). However, subsequent upregulation and distal-anterior propagation of this signal-relay cascade and *Fgf4* expression are dependent on the SHH/FGF4 feedback loop established between the polarizing region and posterior AER (Fig. 5, lower panel). The predominantly nuclear *Formin* proteins¹⁶ act in posterior-distal limb-bud mesenchymal cells responding to the SHH signal (Fig. 1g–i) and are essential to maintain expression of the secondary signal *Gremlin* (Fig. 5, lower panel). Several BMPs are expressed by the limb-bud mesenchyme and/or the AER^{24,25}, but it is unknown at present whether all or only particular BMPs are potential targets of *Gremlin*-mediated antagonism. BMPs and FGFs may antagonize one another during limb-bud development²⁶. Therefore, inhibition of BMP activity in either or both limb-bud compartments by diffusible antagonists such as

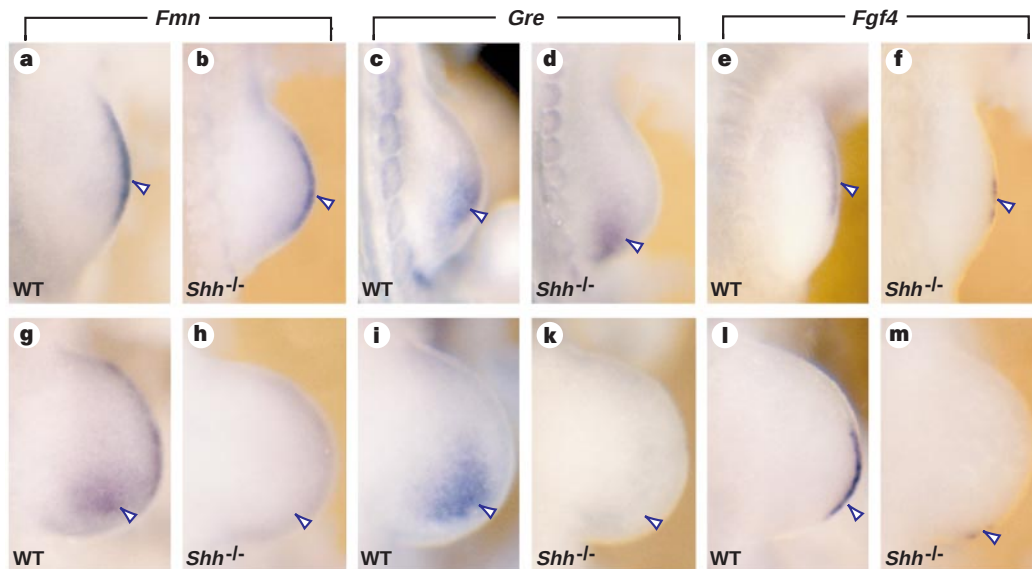


Figure 4 Analysis of *Shh*-deficient mouse limb buds⁹ shows that induction but not maintenance of *Formin*, *Gremlin* and *Fgf4* are SHH independent. **a, b, g, h**, *Formin* (*Fmn*) distribution in wild-type (WT) and *shh*-deficient (*Shh*^{−/−}) hindlimb (**a, b**) and forelimb buds (**g, h**). **c, d, i, k**, *Gremlin* distribution in WT and *Shh*^{−/−} hindlimb (**c, d**) and forelimb buds (**i, k**). **e, f, l, m**, *Fgf4* distribution in WT and *Shh*^{−/−} hindlimb (**e, f**) and forelimb buds

(**l, m**). Note the very small and weak *Fgf4* domain remaining proximal in *Shh*^{−/−} forelimb buds (**m**). All embryos shown are ~E10.25 (32–33 somites). Hindlimb buds lag behind forelimb buds in their development, which enables convenient comparison of 'early' to 'late' distributions in the same embryo. Open arrowheads indicate expression domains in wild-type limb buds and corresponding regions in *Shh*^{−/−} limb buds (for details see text).

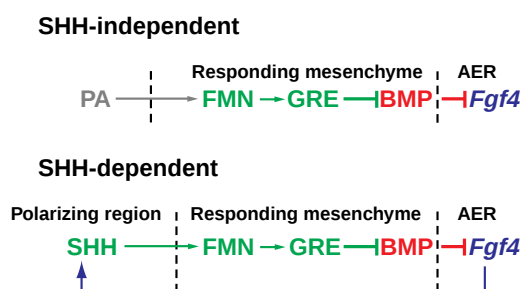


Figure 5 Limb-bud mesenchyme to AER signalling and establishment of the SHH/FGF4 feedback loop through FMN- and GRE-mediated BMP antagonism. Top, SHH-independent induction of the mesenchymal response cascade (*Fmn* and *Gre*) by an unknown posterior activity (PA; see text). Presumed early inhibition of BMP activity by *Gre* allows *Fgf4* activation in the AER independently of SHH signalling. Bottom, SHH-dependent maintenance and distal-anterior propagation of the SHH/FGF-4 feedback loop. In contrast to the activation phase (top), SHH signalling by the polarizing region is essential to maintain and propagate *Fmn* and *Gre* expression in responding mesenchymal cells and *Fgf4* expression in the overlying AER. FGF4 signalling in turn maintains *Shh* expression by the polarizing region, resulting in establishment and maintenance of the SHH/FGF4 feedback loop¹. The genes mediating the initial response to SHH signalling in the limb-bud mesenchyme (*Ptc*, *Smo*, *Gli*) are not indicated³⁰ as they act upstream of *Fmn* (see also Fig. 1d–f). SHH, *Fmn* and *Gre* are shown in green to indicate their positive effect on *Fgf4* expression. BMPs are indicated in red owing to their inhibitory effect on *Fgf4* activation.

Gremlin might override BMP-mediated *Fgf4* repression (Fig. 5; see also ref. 20). In support of this proposal, a *cis*-acting negative regulatory element controlling *Fgf4* expression in the posterior AER has been identified²⁷. This negative regulatory element could be a target of transcriptional repression by BMP signalling. Non-cell-autonomously acting BMP antagonists could locally relieve repression of *Fgf4* activation (Fig. 5), as is expected from its AER domain's being in register with that of *Gremlin* (see Fig. 2j–l). The finding that BMP antagonism is crucial to relay the SHH signal to the AER during vertebrate limb development contrasts sharply with the way the *Hedgehog* (*Hh*) signal is transduced during *Drosophila melanogaster* wing and leg imaginal disc development. Downstream of *Hh* signalling, the BMP homologue *Dpp* functions as a diffusible morphogen, which activates targets in response to specific thresholds²⁸. In contrast to previous work^{1–3,18}, our study does not provide any evidence supporting a similar role for BMP2 during vertebrate limb-bud patterning. Our results indicate that signalling by the vertebrate limb-bud organizer generates differences by extracellular antagonism, rather than only positively acting signals, which reveals molecular similarities to other embryonic organizing centres such as Spemann's organizer. □

Methods

In vitro grafting and culturing of mouse limb buds (trunk cultures)

Isolated mouse embryonic trunk fragments were cultured using an established *in vitro* system^{7,26} with minor modifications²⁹. This culturing method enables *in vitro* development of mouse embryonic limb buds for up to 48 h in serum-free, defined BGJb medium (supplemented with 0.2 mg ml⁻¹ ascorbic acid and penicillin/streptomycin, Sigma) in 5% CO₂ at 37°C. Growth and development of cultured limb buds is only slightly delayed in comparison with *in vivo* development²⁹. Most importantly, the molecular alterations observed in *ld* homozygous limb buds are maintained during culturing. Briefly, either wild-type (+/+) or *ld* mutant (*ld/ld*) mouse embryos (*ld*^{h12} allele)^{10,11} of gestational days 10.25–10.5 (32–35 somites) were isolated, and non-proliferating, spherical cell aggregates (see below) of ~50 µm diameter were grafted into forelimb buds. Following grafting, trunk fragments with attached forelimb buds were isolated and pinned on inverted V-shaped stainless steel grids to allow growth at the medium–air interface²⁹. For this study, culturing of trunk fragments was in general limited to 19–22 h. Only trunks with normally developed limb buds (significant growth and no malformation or AER damage) were included in the analysis. At least 5–10 grafted limb buds from three or more independent experiments were analysed and yielded identical results. No differences were observed using forelimb buds of embryos with 32–35 somites at the time of grafting.

Expressing cells and preparation of cell aggregates

QT6 cells expressing chicken SHH¹³ and Q2bn cells expressing human BMP-2 have been described¹⁸. CHO cells expressing *Xenopus laevis* *Noggin* have been described²³. *Gremlin*-expressing cells: previous studies have successfully investigated the activity of secreted proteins using transient transfection assays (*Gremlin*, SHH)^{4,13}. The murine *Gremlin* coding region⁴ was amplified by polymerase chain reaction with reverse transcription (RT-PCR) and its identity confirmed by sequence analysis. *Gremlin* was expressed in QT6 cells using the RC/CMV vector (Invitrogen) and standard calcium phosphate transfection. About 24 h after transfection, QT6 cells were re-seeded on bacterial dishes to form cell aggregates (see below). Green fluorescent protein (GFP; under control of the CMV promoter pEGFP; Clontech) was used for mock transfection and to monitor transfection efficiency and expression in cell aggregates. *Gremlin* expression was directly assessed by whole mount *in situ* hybridization after culturing, and was detected in all grafts analysed (data not shown). Five transient transfections were included in this analysis and *Fgf4* was detected in all limb buds (Fig. 3). Preparation of spherical cell aggregates: the day before grafting, cells were plated at high density on bacterial dishes to induce formation of cell aggregates. On the day of grafting, cell proliferation was irreversibly blocked by mitomycin-C treatment.

Whole-mount *in situ* hybridization

All analysis was performed as described⁶. Detection of AER transcripts requires proteinase K (Merck) treatment of only 5 µg ml⁻¹ for 4 min, whereas detection of mesenchymal transcripts requires treatment with 10 µg ml⁻¹ proteinase K for 15 min. AER treatment was also used for simultaneous detection of AER and mesenchymal transcripts. Antisense riboprobes were generated from previously described murine templates^{6,12}. We also used an antisense probe complementary to the whole murine *Gremlin* coding region⁴. Finally, we compared the expression of several BMP antagonists in wild-type and *ld* mutant embryos using riboprobes complementary to *Noggin*, *Chordin*, *Cerberus*, *DAN* and *Follistatin*. These genes were not expressed in posterior-distal limb-bud mesenchyme and/or not altered in *ld* mutant limb buds (data not shown). Simultaneous detection of two transcripts using two-colour *in situ* hybridization was achieved by labelling one probe with digoxigenin-UTP and the other with fluorescein-UTP (Boehringer Mannheim). Both probes were hybridized at the same time and the digoxigenin-labelled probe was detected

as usual yielding a purple stain. After fixation and inactivation of the first alkaline phosphatase (AP) complex, the fluoresceinated probe was detected using a INT/BCIP AP-substrate mix (Boehringer Mannheim), which yields an orange-red stain.

Received 5 May; accepted 23 July 1999.

- Johnson, R. L. & Tabin, C. J. Molecular models for vertebrate limb development. *Cell* **90**, 979–990 (1997).
- Lauffer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. & Tabin, C. Sonic hedgehog and Fgf-4 act through a signalling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79**, 993–1003 (1994).
- Niswander, L., Jeffrey, S., Martin, G. R. & Tickle, C. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371**, 609–612 (1994).
- Hsu, D., Economides, A., Wang, X., Eimon, P. & Harland, R. The *Xenopus* dorsalizing factor *Gremlin* identifies a novel family of secreted proteins that antagonise BMP activities. *Mol. Cell* **1**, 673–83 (1998).
- Chan, D. C., Wynshaw-Boris, A. & Leder, P. Formin isoforms are differentially expressed in the mouse embryo and are required for normal expression of *fgf-4* and *shh* in the limb bud. *Development* **121**, 3151–3162 (1995).
- Haramis, A. G., Brown, J. M. & Zeller, R. The limb deformity mutation disrupts the SHH/FGF-4 feedback loop and regulation of 5' HoxD genes during limb pattern formation. *Development* **121**, 4237–4245 (1995).
- Kuhlman, J. & Niswander, L. Limb deformity proteins: role in mesodermal induction of the apical ectodermal ridge. *Development* **124**, 133–139 (1997).
- Chiang, C. et al. Cyclopia and defective axial patterning in mice lacking *Sonic hedgehog* gene function. *Nature* **383**, 407–413 (1996).
- St-Jacques, B. et al. *Sonic hedgehog* signaling is essential for hair development. *Curr. Biol.* **8**, 1058–1068 (1998).
- Woychik, R. P., Maas, R. L., Zeller, R., Vogt, T. F. & Leder, P. 'Formins': Proteins deduced from the alternative transcripts of the limb deformity gene. *Nature* **346**, 850–853 (1990).
- Mass, R. L., Zeller, R., Woychik, R. P., Vogt, T. F. & Leder, P. Disruption of formin-encoding transcripts in two mutant limb deformity alleles. *Nature* **346**, 853–855 (1990).
- Zuniga, A. & Zeller, R. Gli3 (Xt) and formin (ld) participate in the positioning of the polarising region and control of posterior limb-bud identity. *Development* **126**, 13–21 (1999).
- Duprez, D., Fournier-Thibault, C. & LeDouarin, N. *Sonic Hedgehog* induces proliferation of committed skeletal muscle cells in the chick limb. *Development* **125**, 495–505 (1998).
- Marigo, V., Johnson, R. L., Vortkamp, A. & Tabin, C. J. *Sonic hedgehog* differentially regulates expression of *GLI* and *GLI3* during limb development. *Dev. Biol.* **180**, 273–283 (1996).
- Marigo, V., Scott, M. P., Johnson, R. L., Goodrich, L. V. & Tabin, C. J. Conservation in *hedgehog* signaling: induction of a chicken *patched* homolog by *Sonic hedgehog* in the developing limb. *Development* **122**, 1225–1233 (1996).
- Trumpp, A., Blundell, P. A., de la Pompa, J. L. & Zeller, R. The chicken *limb deformity* gene encodes nuclear proteins expressed in specific cell types during morphogenesis. *Genes Dev.* **6**, 14–28 (1992).
- Zeller, R., Jackson-Grusby, L. & Leder, P. The *limb deformity* gene is required for apical ectodermal ridge differentiation and anterior posterior limb pattern formation. *Genes Dev.* **3**, 1481–1492 (1989).
- Duprez, D. M., Kostakopoulou, K., Francis-West, P. H., Tickle, C. & Brickell, P. M. Activation of *Fgf-4* and *HoxD* gene expression by BMP-2 expressing cells in the developing chick limb. *Development* **122**, 1821–1828 (1996).
- Cho, K. & Blitz, I. BMPs, Smads and metalloproteases: extracellular and intracellular modes of negative regulation. *Curr. Opin. Genet. Dev.* **8**, 443–449 (1998).
- Pizette, S. & Niswander, L. BMPs negatively regulate structure and function of the limb apical ectodermal ridge. *Development* **126**, 883–894 (1999).
- McMahon, J. et al. *Noggin*-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes Dev.* **12**, 1438–1452 (1998).
- Brunet, L., McMahon, J., McMahon, A. & Harland, R. *Noggin*, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science* **280**, 1455–1457 (1998).
- Lamb, T. et al. Neural induction by the secreted polypeptide *noggin*. *Science* **262**, 713–718 (1993).
- Francis, P. H., Richardson, M. K., Brickell, P. M. & Tickle, C. Bone morphogenetic proteins and a signalling pathway that controls patterning in the developing chick limb. *Development* **120**, 209–218 (1994).
- Hofmann, C., Luo, G., Balling, R. & Karsenty, G. Analysis of limb patterning in BMP-7-deficient mice. *Dev. Genet.* **19**, 43–50 (1996).
- Niswander, L. & Martin, G. R. FGF-4 and BMP-2 have opposite effects on limb growth. *Nature* **361**, 68–71 (1993).
- Fraidenraich, D., Lang, R. & Basilio, C. Distinct regulatory elements govern *Fgf4* gene expression in the mouse blastocyst, myotome, and developing limb. *Dev. Biol.* **204**, 197–209 (1998).
- Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range action of a DPP morphogen gradient. *Cell* **85**, 357–368 (1996).
- Lussier, M., Canoun, C., Ma, C., Sank, A. & Shuler, C. Interdigital soft tissue separation induced by retinoic acid in mouse limbs cultured *in vitro*. *Int. J. Dev. Biol.* **37**, 555–564 (1993).
- Ingham, P. Transducing *Hedgehog*: the story so far. *EMBO J.* **17**, 3505–3511 (1998).

Acknowledgements

We thank G. Drossopoulou, M. Herlevsen and C. Tickle for advice in establishing the *in vitro* culturing system; T. Bouwmeester and S. Cohen for many stimulating discussions and support; C. McGuigan, B. Wagemakers and H. Goedemans for technical assistance; J. McMahon for breeding *Shh*-deficient mice; J. Jackson for mouse husbandry; D. Duprez, P. Brickell, H. Rohrer, R. Harland, T. Bouwmeester, S.-L. Ang and C. Niehrs for reagents; and G. Davidson, R. Dono, A. van Loon, P. Rorth and L. Panman for comments on the manuscript. This study was supported by EMBL and Utrecht University, grants from the European Community (to R.Z.), the NIH (to A.P.M.) and the Graduiertenkolleg "Experimentelle Nieren- und Kreislaufforschung" (to A.Z.).

Correspondence and requests for materials should be addressed to R.Z. at Utrecht University (e-mail: R.Zeller@bio.uu.nl).

Long-term *in vivo* reconstitution of T-cell development by Pax5-deficient B-cell progenitors

Antonius G. Rolink*, Stephen L. Nutt†‡, Fritz Melchers* & Meinrad Busslinger†

* Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland

† Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria

The mechanisms controlling the commitment of haematopoietic progenitors to the B-lymphoid lineage are poorly understood. The observations that mice deficient in E2A^{1,2} and EBF³ lack B-lineage cells have implicated these two transcription factors in the commitment process. Moreover, the expression of genes encoding components of the rearrangement machinery (RAG1, RAG2, TdT) or pre-B-cell receptor (λ 5, VpreB, Ig α , Ig β) has been considered to indicate B-lineage commitment⁴. All these genes including E2A and EBF are expressed in pro-B cells lacking the transcription factor Pax5 (refs 5–7). Here we show that cloned Pax5-deficient pro-B cells transferred into RAG2-deficient mice provide long-term reconstitution of the thymus and give rise to mature T cells expressing α/β -T-cell receptors. The bone marrow of these mice contains a population of cells of Pax5^{-/-} origin with the same phenotype as the donor pro-B cells. When transferred into secondary recipients, these pro-B cells again home to the bone marrow and reconstitute the thymus. Hence, B-lineage commitment is determined neither by immunoglobulin DJ rearrangement nor by the expression of E2A, EBF, λ 5, VpreB, Ig α and Ig β . Instead, our data implicate Pax5 in the control of B-lineage commitment.

The B-cell progenitors known as pre-BI cells have been defined by their expression of B220, c-Kit, λ 5 and VpreB, the rearrangement status of their immunoglobulin heavy-chain (IgH) loci and their *in vitro* growth capacity in the presence of stromal cells and IL-7 (refs 8, 9). The bone marrow of Pax5^{-/-} mice also contains cells with a similar surface phenotype and *in vitro* growth property⁶. In contrast to pre-BI cells from wild-type bone marrow, the Pax5^{-/-} pre-BI cells do not lose their growth capacity with time⁶. For this and other reasons mentioned below, we refer here to the Pax5^{-/-} pre-BI cells as pro-B cells.

Transplantation experiments in RAG-deficient mice revealed that *in vitro* cultured wild-type pro-B cells can differentiate *in vivo* only along the B-cell pathway^{10,11}. As Pax5^{-/-} pro-B cells can differentiate along myeloid lineages *in vitro*¹², we tested their *in vivo* developmental potential by injecting 5–10 $\times$ 10⁶ Pax5^{-/-} pro-B cells (H-2^b haplotype) into sublethally irradiated BALB/c RAG2^{-/-} mice (H-2^d haplotype). Alternatively, 5–10 $\times$ 10⁶ Pax5^{-/-} pro-B cells, infected with a green fluorescent protein (GFP)-expressing retrovirus, were injected into irradiated syngeneic C57BL/6 RAG2^{-/-} mice. Both experiments gave similar results. The largest number of cells derived from the injected Pax5^{-/-} pro-B cells were found in the thymus (Fig. 1b). By three weeks after transplantation, 5 $\times$ 10⁷–10⁸ cells of donor origin were present in the thymus of the recipients. Flow cytometric (FACS) analysis revealed that the Pax5^{-/-} pro-B cells reconstituted the thymus of RAG2^{-/-} hosts with all the thymocyte subpopulations in ratios identical to those found in thymus of wild-type mice, including CD4⁺CD8⁻ (double-negative), CD4⁺CD8⁺ (double-positive) and CD4⁺CD8⁻ and CD8⁺CD4⁻ (single-positive) cells (Fig. 1b). Moreover, these thymocyte subpopulations expressed

levels of α/β -T-cell receptors (TCR) similar to those of wild-type thymocytes, with high levels on single-positive cells and intermediate and low levels on double-positive populations. T cells expressing γ/δ -TCR were also found in these reconstituted thymus at the frequency (0.5–1%) seen in CD4⁺CD8⁻ cells of wild-type thymus.

Pax5^{-/-} pro-B cells, like their wild-type counterparts, transcribe λ 5 and VpreB but do not express pT α , CD3 δ or any other CD3 gene (Fig. 2a and data not shown). However, all thymocyte subpopulations in reconstituted RAG2^{-/-} mice lost λ 5 and VpreB expression and now expressed CD3 δ (Fig. 2a). As in the thymus of wild-type mice¹³, pT α expression was restricted to the double-negative and double-positive stages of development. Moreover, all thymocyte subpopulations retained the D_HJ_H-rearranged IgH loci present in the pro-B cells and additionally carried D β 2J β 2-rearranged TCR β loci (Fig. 2b).

Figure 3 shows the time course of thymic reconstitution after transplantation of GFP-expressing Pax5^{-/-} pro-B cells. At day 7 after injection, GFP-expressing cells were rare and scattered throughout the thymus (Fig. 3b). By day 11, the cortex of the thymus was reconstituted with GFP-expressing cells, whereas few fluorescent cells were present in the medulla (Fig. 3c). By day 16, the whole thymus contained GFP-expressing cells (Fig. 3d). This temporal and spatial pattern of thymus reconstitution is very similar to that observed in lethally irradiated mice after bone-marrow transfer¹⁴. Thymus reconstitution was still observed 4 months after transplantation. From week 3 onwards, mature single-positive, α/β -TCR-expressing T-cells were also found in spleen and lymph nodes (Fig. 1a). However, mature B-lymphocytes were never observed in the peripheral lymphoid organs of reconstituted mice (unpublished data).

Next we investigated whether the reconstitution of T-cell development is a clonal property of the Pax5^{-/-} pro-B cell. Twelve different pro-B-cell clones with characteristic D_HJ_H rearrangements at the IgH locus were established by single-cell sorting. Upon transfer into RAG2^{-/-} mice, eight clones gave rise to thymus reconstitution. Characterization of the IgH locus in subpopulations of the reconstituted thymus showed that they carried the same D_HJ_H rearrangements as the original pro-B-cell clones. A representative polymerase chain reaction (PCR) analysis is shown for the Pax5^{-/-} pro-B-cell clone K2.11, which contains a D_HJ_H1- and D_HJ_H3-rearranged IgH allele (Fig. 2c). This result rules out the possibility that a rare cell different from the Pax5^{-/-} pro-B cell is responsible for thymus reconstitution.

It might be argued that the long-term *in vitro* growth could change the Pax5-deficient pro-B cells such that they acquired the capacity to colonize the thymus of RAG2^{-/-} mice. To exclude this possibility, 10⁴ B220⁺c-Kit⁺ pro-B cells were sorted from Pax5^{-/-} bone marrow and directly injected into irradiated RAG2^{-/-} mice. Three weeks after transfer, the thymus of these mice contained 5–7 $\times$ 10⁷ donor-derived cells. FACS analysis revealed normal ratios of CD4CD8 double-negative, double-positive and single-positive cell populations, but no B-cells (Fig. 1c). Thus, the *ex vivo* isolated Pax5^{-/-} pro-B cells possess the same properties as their *in vitro* grown counterparts with the exception that at least 100-fold fewer *ex vivo* sorted cells are required for thymus reconstitution.

As Pax5^{-/-} pro-B cells can reconstitute T-cell development in RAG2^{-/-} mice, we investigated whether a cell with similar properties could be found in the bone marrow of wild-type mice. In support of such a cell, D_HJ_H rearrangements at the IgH locus¹⁵ and expression of Ig β ¹⁶ have been observed in thymocytes. Moreover, a clonogenic common lymphoid progenitor (CLP) with B-, T- and natural killer (NK)-cell potential was isolated from bone marrow¹⁷. Pro-B cells from wild-type mice, in contrast to those from Pax5^{-/-} mice, give rise only to donor-derived B cells, but not to T cells, in transplanted hosts^{10,11}. As the Pax5^{-/-} pro-B cells, unlike wild-type pro-B cells, do not express CD19 (refs 6, 7), we sorted CD19⁺B220⁺c-Kit⁺ cells (0.1–0.2% of total nucleated cells) from wild-type bone marrow

‡ Present address: Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, UK.

and transplanted these cells into irradiated $RAG2^{-/-}$ mice. Three weeks after transplantation of 10^4 cells, spleen and thymus were analysed. A small but significant number of mature B cells were found, which persisted for up to 3 months (Fig. 1d). We saw neither reconstitution of the thymus nor T-cell development in the periphery (Fig. 1d). Thus, a cell with the $CD19^+B220^+c\text{-Kit}^+$ phenotype cannot be the normal counterpart of the $Pax5^{-/-}$ pro-B cell. Expression of the interleukin-7 receptor (IL-7R) has been used as a criterion for isolating the CLP¹⁷. $Pax5^{-/-}$ pro-B cells also express IL-7R and require IL-7 for proliferation⁶, indicating a close relationship between this cell and the CLP. However, the CLP cannot be propagated long-term under *in vitro* IL-7 growth conditions without losing its T- and NK-cell potential¹⁷. Instead, these cells become committed to and differentiate along the B-cell pathway¹⁷. Under similar growth conditions, the $Pax5^{-/-}$ pro-B cells cannot undergo B-lineage commitment and thus retain their T-cell potential.

Committed T-lymphoid progenitors give rise to only one wave of T-cell development lasting for about two weeks¹⁸. In contrast, $Pax5^{-/-}$ pro-B cells fully reconstitute T-cell development in $RAG2^{-/-}$ mice for at least 4 months. As the thymus is continuously seeded by bone marrow-derived progenitors, we analysed $RAG2^{-/-}$ mice injected with GFP-expressing $Pax5^{-/-}$ pro-B cells for the

presence of fluorescent cells in the bone marrow. Indeed, ~5% of the bone marrow cells were GFP⁺, expressed B220, c-Kit, $\lambda 5$ and VpreB and carried the $IgH D_{HJH}$ rearrangements characteristic of the transferred $Pax5^{-/-}$ pro-B cells (Fig. 4a and data not shown). Hence, the injected $Pax5^{-/-}$ cells homed to the bone marrow where they maintained their pro-B-cell phenotype. When these cells were isolated, they grew efficiently on stromal cells with IL-7. Transplantation of these cells into secondary $RAG2^{-/-}$ mice again resulted in full reconstitution of the thymus. Moreover, the bone marrow of these secondary recipients also contained GFP⁺B220⁺c-Kit⁺ cells of $Pax5^{-/-}$ origin (Fig. 4b). Analysis of DNA content revealed that 7.5% of these cells were in the S, G2 or M phase of the cell cycle (Fig. 4b). Thus, *in vitro* cultured $Pax5^{-/-}$ pro-B cells have the capacity for self-renewal in the bone marrow environment of primary and secondary $RAG2^{-/-}$ recipients, explaining their long-term thymus-reconstitution potential.

The question therefore arises of whether the $Pax5^{-/-}$ pro-B cells in the bone marrow can also generate haematopoietic cell types other than T cells. Upon *in vitro* stimulation with appropriate cytokines, $Pax5^{-/-}$ pro-B cells can indeed give rise to macrophages, osteoclasts, dendritic cells, granulocytes and NK cells¹². However, we have not detected the formation of these cell types in numbers detectable by

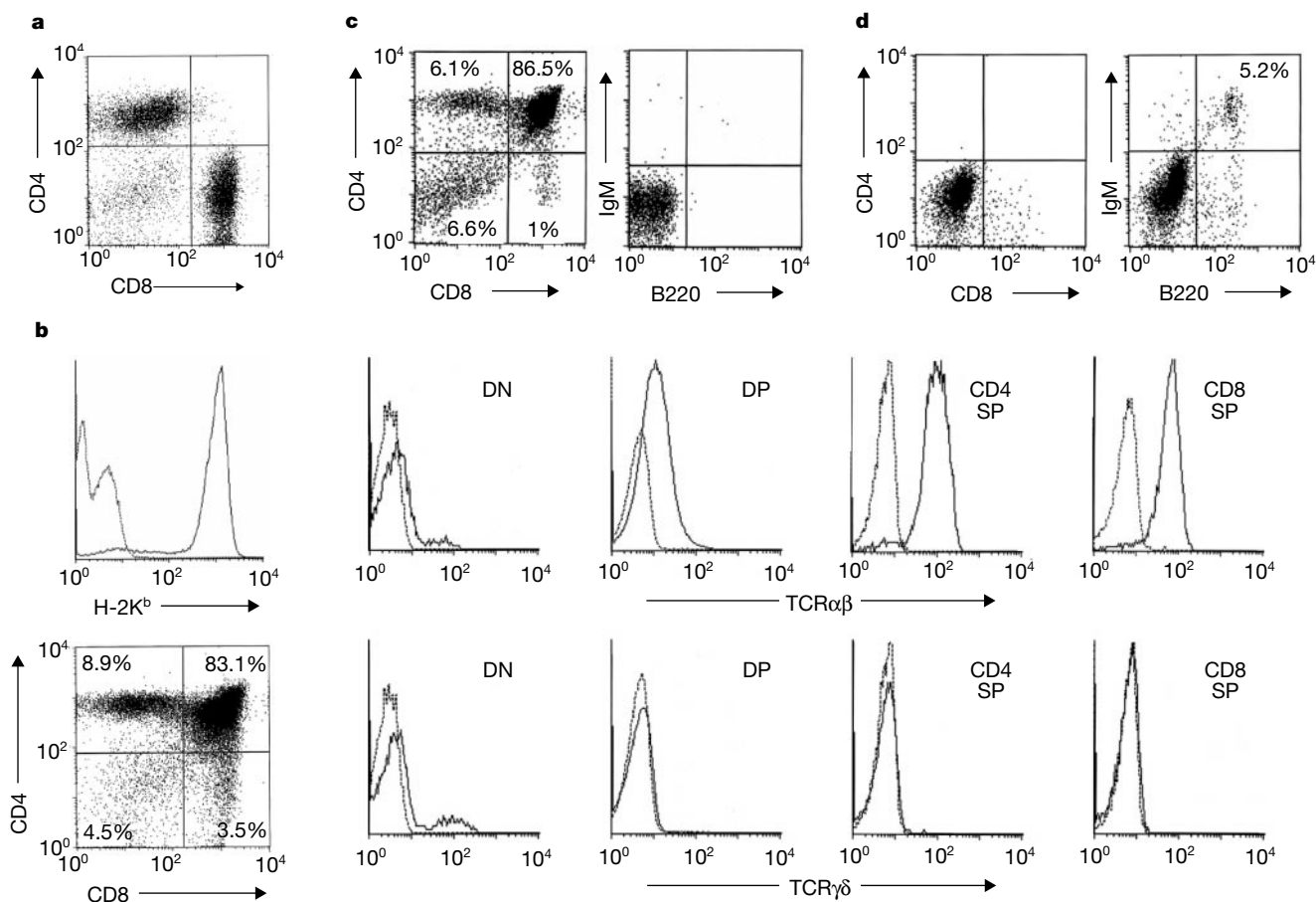


Figure 1 Reconstitution of T-cell development by $Pax5^{-/-}$ pro-B cells. **a**, Appearance of mature $CD4^+$ and $CD8^+$ T-cells in the lymph node of $RAG2^{-/-}$ mice reconstituted with 10^7 $Pax5^{-/-}$ pro-B cells. At day 28 after transplantation, FACS analysis was performed with FITC-anti-CD4 and PE-anti-CD8 antibodies. **b**, Thymus reconstitution by $Pax5^{-/-}$ pro-B cells. 10^7 $Pax5^{-/-}$ pro-B cells of the H-2^b haplotype were injected into 400 rad-irradiated BALB/c $RAG2^{-/-}$ mice. At day 18 after transfer, the thymus was analysed by flow cytometry, indicating that >95% of all thymocytes were H-2^b positive and thus of $Pax5^{-/-}$ origin (upper row, left). Three-colour FACS analysis using PE-anti-CD4, APC-anti-CD8 and FITC-anti-TCR β or FITC-anti-TCR $\gamma\delta$ antibodies revealed $CD4^+CD8^-$ double-negative (DN), $CD4^+CD8^+$ double-positive (DP), $CD4^+$ and $CD8^+$ single-positive (SP) cells

at normal ratios (lower row, left). These subpopulations (gated on $CD4^+CD8^+$ expression) exhibited similar α/β and γ/δ -TCR expression to the respective wild-type thymocytes. Unstained control cells are indicated by dashed lines. **c**, $CD19^+B220^+c\text{-Kit}^+$ pro-B cells isolated *ex vivo* from $Pax5^{-/-}$ bone marrow also reconstitute the thymus of $RAG2^{-/-}$ mice, but do not generate sIgM⁺ B-cells. **d**, $CD19^+B220^+c\text{-Kit}^+$ cells isolated from wild-type mice fail to reconstitute the thymus, but generate sIgM⁺ B-cells. $CD19^+B220^+c\text{-Kit}^+$ cells sorted from $Pax5^{-/-}$ or wild-type bone marrow were injected into 400 rad-irradiated $RAG2^{-/-}$ mice. After 3 weeks, the thymus and spleen were analysed by flow cytometry, using PE-anti-CD4 and FITC-anti-CD8 antibodies (thymocytes) or PE-anti-IgM and FITC-anti-B220 antibodies (splenocytes), respectively.

FACS analysis in reconstituted $RAG2^{-/-}$ mice. As the $RAG2^{-/-}$ mice have no defect in these lineages, it is conceivable that the $Pax5^{-/-}$ pro-B cells cannot compete with endogenous haematopoietic progenitors. Therefore, the *in vivo* myeloid potential of $Pax5^{-/-}$ pro-B cells may be better studied in mouse strains which are genetically defective in one of these lineages. Indeed, $Pax5^{-/-}$ pro-B cells can generate osteoclasts¹² in osteoclast-deficient $c-fos^{-/-}$ mice¹⁹.

$Pax5^{-/-}$ pro-B cells express the transcription factors E2A and EBF as well as the early B-lymphoid genes $\lambda 5$, $VpreB$, $Ig\alpha$ and $Ig\beta$, which have been considered to indicate B-lineage commitment⁷. Our results show that these genes are not directly involved in commitment to the B lineage. Instead, this process critically depends on the transcription factor Pax5 which appears to exert a dual function at B-lineage commitment by activating B-lymphoid genes⁷ and simultaneously repressing lineage-inappropriate genes¹². The availability of $Pax5^{-/-}$ pro-B cells will now facilitate the search for these target genes. □

Methods

Mice

$RAG2$ -deficient mice²⁰ were originally provided by F. Alt. BALB/c $RAG2^{-/-}$ and C57BL/6 $RAG2^{-/-}$ mice were generated by back-crossing the $RAG2$ mutation into BALB/c or C57BL/6 mice for 15 generations and were kept under special pathogen-free conditions.

Cell lines

Pro-B cell lines from $Pax5^{-/-}$ mice⁵ were grown as described^{6,8}. Clones of pro-B cells were generated by sorting single cells from established pro-B cell lines, using the FACStar Plus equipped with the automatic cell-deposit unit (Becton Dickinson). Thereafter, the single cells were grown into clones on stromal cells in the presence of IL-7. GFP-expressing pro-B

cells were generated by retroviral infection, by coculturing pro-B cells overnight in the presence of IL-7 with GP + E-86 packaging cells²¹ which had been transfected with the LXS-puro vector²² expressing GFP. Complementary DNA coding for enhanced GFP was obtained from Life Technologies. This coculture produces 30–70% GFP-expressing pro-B cells. After another 6 days of puromycin selection ($2.5 \mu\text{g ml}^{-1}$, Sigma) on puromycin-resistant stromal cells in the presence of IL-7, nearly all pro-B cells expressed high levels of GFP.

Transfer of cells

C57BL/6 $RAG2^{-/-}$ or BALB/c $RAG2^{-/-}$ mice at 8–12 weeks of age were γ -irradiated with 400 rad and injected 6–8 h after irradiation i.v. with $5\text{--}10 \times 10^6$ *in vitro* grown $Pax5^{-/-}$ pro-B cells. In a second experimental set-up, irradiated $RAG2^{-/-}$ mice were injected i.v. with 10^4 B220⁺c-Kit⁺CD19⁺ cells derived from the bone marrow of normal C57BL/6 mice or $Pax5^{-/-}$ mice. Cells were purified by sorting in a FACStar Plus flow cytometer (Becton Dickinson). Injected $RAG2^{-/-}$ mice were treated with neomycin sulphate (Sigma) which was added to the drinking water at a final concentration of 1.17 g l^{-1} . This antibiotic treatment prevented opportunistic infections which otherwise resulted in fulminant immune responses, killing the recipient mouse ~6 weeks after transplantation.

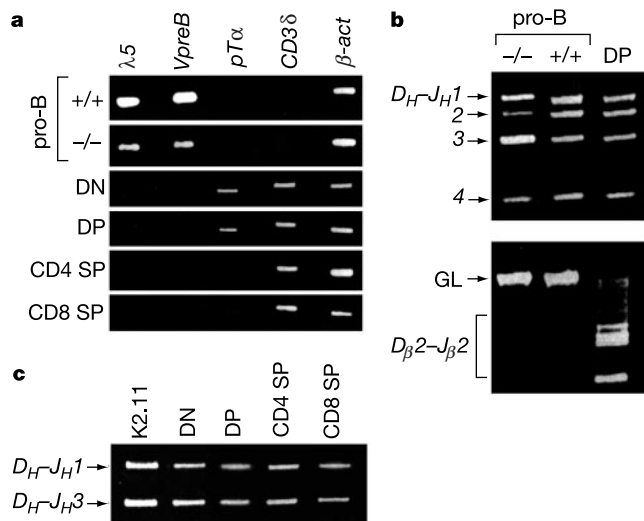


Figure 2 Gene expression and rearrangement in thymocytes derived from $Pax5^{-/-}$ pro-B cells. **a**, RT-PCR analysis. Expression of $\lambda 5$, $VpreB$, $pT\alpha$, $CD3\delta$ and of the control β -actin (β -act) gene was analysed by RT-PCR in *in vitro* cultured wild-type (+/+) and $Pax5$ -deficient (-/-) pro-B cells as well as in $CD4^{-}CD8^{-}$ (DN), $CD4^{+}CD8^{+}$ (DP), and $CD4^{+}$ and $CD8^{+}$ (SP) thymocytes of reconstituted $RAG2^{-/-}$ mice. These thymocyte subpopulations were sorted from BALB/c $RAG2^{-/-}$ mice 20 days after transplantation with 10^7 $Pax5^{-/-}$ pro-B cells. **b**, Gene rearrangements. $Pax5$ -deficient (-/-) and wild-type (+/+) pro-B cells and $CD4^{+}CD8^{+}$ (DP) thymocytes, which were sorted from $RAG2^{-/-}$ mice reconstituted with $Pax5^{-/-}$ pro-B cells, were analysed by PCR for *DJ* rearrangements at the *IgH* and *TCRB* loci. The sizes of the various D_HJ_H and $D_\beta J_\beta$ PCR fragments have been described^{16,18,24}. The position of the germline (GL) PCR fragment is indicated. **c**, T-cell reconstitution by the $Pax5^{-/-}$ pro-B cell clone K2.11 which was generated by single-cell sorting. This clone carries a D_HJ_H1 - and D_HJ_H3 -rearranged *IgH* allele. Three weeks after transfer of 10^7 K2.11 cells into $RAG2^{-/-}$ mice, $CD4^{-}CD8^{-}$ (DN), $CD4^{+}CD8^{+}$ (DP), and $CD4^{+}$ and $CD8^{+}$ (SP) thymocytes were isolated by sorting and analysed for *D_HJ_H* rearrangements by PCR.

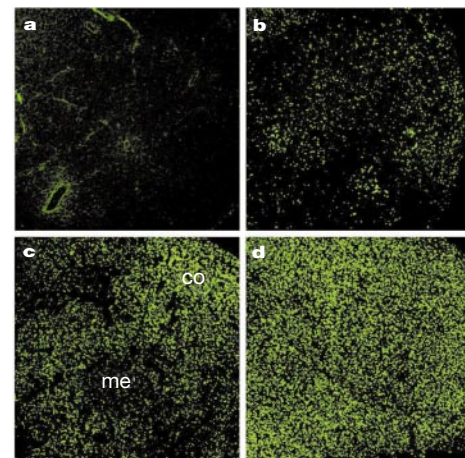


Figure 3 Time course of thymus reconstitution. GFP-expressing cells were detected in the thymus of $RAG2^{-/-}$ mice at day 0 (**a**), day 7 (**b**), day 11 (**c**) and day 16 (**d**) after transfer of 10^7 GFP-expressing $Pax5^{-/-}$ pro-B cells. me, medulla; co, cortex.

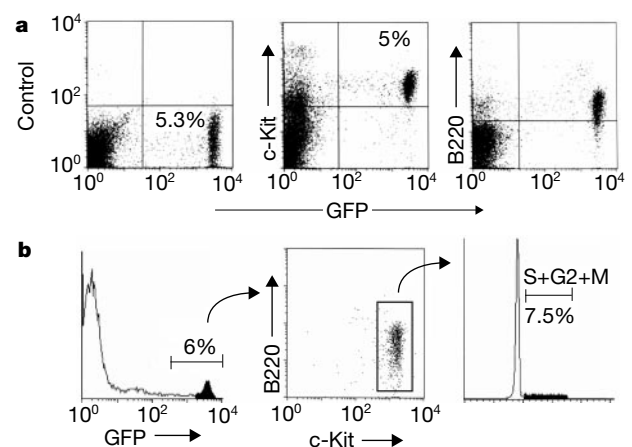


Figure 4 Homing and self-renewal of $Pax5^{-/-}$ pro-B cells in the bone marrow. **a**, Primary transfer. Cells of the GFP-expressing $Pax5^{-/-}$ pro-B cell clone B3 home to the bone marrow of $RAG2^{-/-}$ mice and maintain their pro-B cell phenotype. Four weeks after transfer of 10^7 cells, the bone marrow of $RAG2^{-/-}$ recipients was analysed by flow cytometry for B220 and c-Kit expression. Control refers to unstained cells. **b**, Secondary transfer. GFP⁺B220⁺c-Kit⁺ cells sorted from the primary $RAG2^{-/-}$ recipient were expanded *in vitro* in stromal cells with IL-7 before injection into a secondary $RAG2^{-/-}$ host. Four weeks after transfer, GFP⁺B220⁺c-Kit⁺ cells were sorted from the bone marrow and subjected to cell cycle analysis⁹.

Antibodies and staining

Flow cytometry was performed with a FACSCalibur (Becton Dickinson) as described^{15,23}. Phycoerythrin (PE)- or allophycocyanin (APC)-labelled anti-CD4 (L3T4), anti-CD8 (53-6.7) and anti-B220 (RA3-6B2), PE- or fluorescein isothiocyanate (FITC)-labelled anti-TCR β (H57-597) and anti-IgM (M41), FITC-labelled anti-TCR $\gamma\delta$ (GL3) and PE-labelled anti-H-2K^b (AF6-88.5) antibodies were purchased from PharMingen. Cell-cycle analysis was done by flow cytometry as described⁹.

PCR analysis

The PCR conditions and primers for amplifying *D_HH_H* rearrangements were as described^{6,24}. *D_β2/β2* rearrangements of the *TCRβ* locus were determined, using known PCR conditions and primers recognizing sequences 5' of the *Dβ2.1* element or 3' of the *Jβ2.7* element¹⁸. The expression of the *λ5*, *VpreB*, *pTα* and *β-actin* genes was determined by PCR with reverse transcription (RT-PCR), using published primers and conditions^{25,26}.

GFP analysis in thymus cryosections

To preserve GFP expression, thymi were fixed overnight in 4% paraformaldehyde at 4 °C. Fixed thymi were embedded in Tissue-Tek OCT compound, and 6 μm-thick cryosections were prepared. After mounting in PBS containing 50% glycerol, sections were analysed on a fluorescence microscope (Zeiss Axiophot) equipped with a CCD camera (Photometrics).

Received 12 May; accepted 9 August 1999.

- Bain, G. *et al.* E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* **79**, 885–892 (1994).
- Zhuang, Y., Soriano, P. & Weintraub, H. The helix-loop-helix gene E2A is required for B cells formation. *Cell* **79**, 875–884 (1994).
- Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**, 263–267 (1995).
- Melchers, F. & Rolink, A. in *Fundamental Immunology* (ed. Paul, W. E.) 183–224 (Lippincott-Raven, Philadelphia, 1999).
- Urbánek, P., Wang, Z.-Q., Fetka, I., Wagner, E. F. & Busslinger, M. Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79**, 901–912 (1994).
- Nutt, S. L., Urbánek, P., Rolink, A. & Busslinger, M. Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the *IgH* locus. *Genes Dev.* **11**, 476–491 (1997).
- Nutt, S. L., Morrison, A. M., Dörfler, P., Rolink, A. & Busslinger, M. Identification of BSAP (Pax5) target genes in early B-cell development by loss- and gain-of-function experiments. *EMBO J.* **17**, 2319–2333 (1998).
- Rolink, A., Kudo, A., Karasuyama, H., Kikuchi, Y. & Melchers, F. Long-term proliferating early pre B cell lines and clones with the potential to develop to surface-Ig positive, mitogen-reactive B cells *in vitro* and *in vivo*. *EMBO J.* **10**, 327–336 (1991).
- Rolink, A., Grawunder, U., Winkler, T. H., Karasuyama, H. & Melchers, F. IL-2 receptor α chain (CD25, TAC) expression defines a crucial stage in pre-B cell development. *Int. Immunol.* **6**, 1257–1264 (1994).
- Reininger, L., Radaszkiewicz, T., Kosco, M., Melchers, F. & Rolink, A. G. Development of autoimmune disease in SCID mice populated with long-term *in vitro* proliferating (NZB $\times$ NZW)F1 pre-B cells. *J. Exp. Med.* **176**, 1343–1353 (1992).
- Reininger, L. *et al.* Intrinsic B cell defects in NZB and NZW mice contribute to systemic lupus erythematosus in (NZB $\times$ NZW)F1 mice. *J. Exp. Med.* **184**, 853–861 (1996).
- Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556–562 (1999).
- Fehling, H. J., Krotkova, A., Saint-Ruf, C. & von Boehmer, H. Crucial role of the pre-T-cell receptor- α gene in development of α/β but not γ/δ T cells. *Nature* **375**, 795–798 (1995).
- Ceredig, R. & Schreyer, M. Immunohistochemical localization of host and donor-derived cells in the regenerating thymus of radiation bone marrow chimeras. *Thymus* **6**, 15–26 (1984).
- Born, W., White, J., Kappler, J. & Marrack, P. Rearrangement of *IgH* genes in normal thymocyte development. *J. Immunol.* **140**, 3228–3232 (1988).
- Wang, H., Diamond, R. A. & Rothenberg, E. V. Cross-lineage expression of Ig- β (B29) in thymocytes: positive and negative gene regulation to establish T cell identity. *Proc. Natl Acad. Sci. USA* **95**, 6831–6836 (1998).
- Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
- Rodewald, H.-R., Kretzschmar, K., Takeda, S., Hohl, C. & Dessing, M. Identification of pro-thymocytes in murine fetal blood: T lineage commitment can precede thymus colonisation. *EMBO J.* **13**, 4229–4240 (1994).
- Grigoriadis, A. *et al.* Fos: a key regulator of osteoclast-macrophage lineage determination and bone remodeling. *Science* **266**, 443–448 (1994).
- Shinkai, Y. *et al.* RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855–867 (1992).
- Markowitz, D., Goff, S. & Bank, A. A safe packaging line for gene transfer: separating viral genes on two different plasmids. *J. Virol.* **62**, 1120–1124 (1988).
- Backstrom, B. T. *et al.* A motif within the T cell receptor α chain constant region connecting peptide domain controls antigen responsiveness. *Immunity* **5**, 437–447 (1996).
- Rolink, A., Grawunder, U., Haasner, D., Strasser, A. & Melchers, F. Immature surface Ig⁺ B cells can continue to rearrange κ and λ L chain gene loci. *J. Exp. Med.* **178**, 1263–1270 (1993).
- Gu, H., Kitamura, D. & Rajewsky, K. B cell development regulated by gene rearrangement: arrest of maturation by membrane-bound D μ protein and selection of D H element rearing frames. *Cell* **65**, 47–54 (1991).
- Grawunder, U., Haasner, D., Melchers, F. & Rolink, A. Rearrangement and expression of κ light chain genes can occur without μ heavy chain expression during differentiation of pre-B cells. *Int. Immunol.* **5**, 1609–1618 (1993).

26. Bruno, L., Rocha, B., Rolink, A., von Boehmer, H. & Rodewald, H. R. Intra- and extra-thymic expression of the pre-T cell receptor α gene. *Eur. J. Immunol.* **25**, 1877–1882 (1995).

Acknowledgements

We thank J. Andersson, K. Karjalainen and E. Palmer for critical reading of the manuscript; A. Groenewegen, N. Straube, M. Dessing and A. Pickert for technical assistance; E. Wagner for animal husbandry; and T. Winkler for the GFP retrovirus. This work was supported by the Basel Institute for Immunology, the I.M.P. Vienna and a grant from the Austrian Industrial Research Promotion Fund. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche Ltd., Basel, Switzerland.

Correspondence and requests for material should be addressed to A.G.R. (e-mail: rolink@bii.ch).

Dimerization inhibits the activity of receptor-like protein-tyrosine phosphatase- α

Guoqiang Jiang^{*}, Jeroen den Hertog[†], Jing Su[§], Joseph Noel[‡], Jan Sap[§] & Tony Hunter^{*}

^{*} Molecular Biology and Virology Laboratory and [‡] Structural Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Rd, La Jolla, California 92037, USA

[†] Hubrecht Laboratory, Netherlands Institute for Developmental Biology, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

[§] New York University Medical Center, Department of Pharmacology, New York, New York 10016, USA

Protein-tyrosine phosphatases (PTPs) are vital for regulating tyrosine phosphorylation in many processes, including growth and differentiation^{1,2}. The regulation of receptor-like PTP (RPTP) activity remains poorly understood, but based on the crystal structure of RPTP α domain 1 we have proposed that dimerization can negatively regulate activity, through the interaction of an inhibitory 'wedge' on one monomer with the catalytic cleft of domain 1 in the other monomer^{3,4}. Here we show that dimerization inhibits the activity of a full-length RPTP *in vivo*. We generated stable disulphide-bonded full-length RPTP α homodimers by expressing mutants with single cysteines at different positions in the ectodomain juxtamembrane region. Expression of wild-type RPTP α and Phe135Cys and Thr141Cys mutants in RPTP α -null mouse embryo cells increased dephosphorylation and activity of Tyr 529 in the protein tyrosine kinase c-Src; in contrast, expression of a Pro137Cys mutant did not. Mutation of Pro 210/211 to leucine in the inhibitory wedge of the Pro137Cys mutant restored its ability to activate c-Src, indicating that dimerization may inhibit full-length RPTP α activity in a manner stereochemically consistent with RPTP α crystal structures³. Our results suggest that RPTP α activity can in principle be negatively regulated by dimerization *in vivo*.

Of the ~75 PTP members identified so far, ~25 are RPTPs². The mechanisms for regulating RPTP activity are mostly unknown, but emerging functional evidence indicates that homodimerization may inhibit CD45 RPTP activity^{5–7}, and heterodimerization inhibits RPTP σ (ref. 8). RPTP α has a typical RPTP structure with two cytoplasmic catalytic domains (Fig. 1a)⁹, which are both active¹⁰. RPTP α associates with c-Src¹¹ and directly dephosphorylates the negative regulatory Tyr in c-Src (Tyr 529 in mouse c-Src), leading to its activation^{12–15}. In two crystal forms, murine RPTP α -D1 exists as a symmetrical dimer, in which a wedge-shaped helix–turn–helix element on one monomer inserts into the active site of the dyad-related monomer, resulting in mutual active-site occlusion. This

indicated that RPTP α may be able to form dimers and that, if formed, these dimers would have decreased PTP activity³.

To test this model, we generated covalently stabilized RPTP α dimers with an intermolecular disulphide bond by introducing single unpaired Cys at several positions in the extracellular domain in the region immediately adjacent to the transmembrane domain (TMD) (residues 143–166), as done for Neu/ErbB2 (refs 16, 17) and the epidermal growth factor (EGF) receptor¹⁸. To maximize the chances of obtaining disulphide-linked RPTP α dimers, we made four single Cys mutations (F135C, P137C, D139C and T141C) within a seven-residue span (residues 135–141), directly amino-terminal to the TMD (residues 143–166) of a haemagglutinin (HA)-tagged full-length (FL) RPTP α construct (Fig. 1a). As there

are no other Cys in the RPTP α ectodomain, the free Cys should form an interchain disulphide bond between two RPTP α molecules if they can dimerize, thus stabilizing the dimer. Because TMDs are generally α -helical and are thought to emerge from the membrane in an α -helical conformation, these mutants would place a Cys at four different phases on the α -helical wheel, which repeats every seven residues. Dimers of the four Cys mutants should have the two intracellular catalytic domains in different relative rotational orientations, which could be important if a specific spatial relationship is necessary to bring about attenuation of PTP activity by dimerization, as is the case for Neu/ErbB2 activation¹⁷.

The wild-type RPTP α and Cys-mutant proteins, which were overexpressed in HEK 293 cells by transient transfection, all

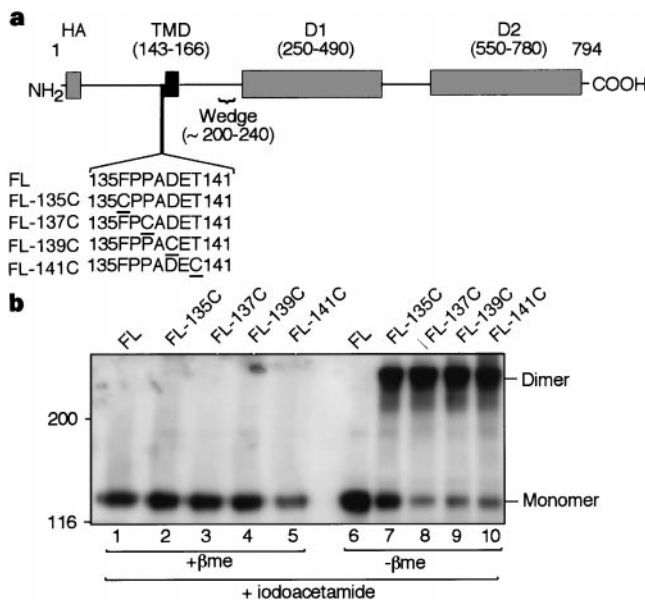


Figure 1 Dimerization of full-length RPTP α via intermolecular disulphide bonds.

a, Diagram of HA-tagged full-length wild type (FL) RPTP α and Cys mutants based on the original (untagged) polypeptide⁹ and D1 crystal structure³. **b**, Reducing and nonreducing SDS-PAGE and immunoblotting (IB) analysis of the whole cell lysates of transiently transfected HEK 293 cells. + iodoacetamide, 20 mM iodoacetamide in lysis buffer; + or - β me, presence or absence of β -mercaptoethanol in the loading buffer. The positions of relative molecular mass markers (in thousands) are indicated on the left side of the gel. Similar abbreviations are used in all figures.

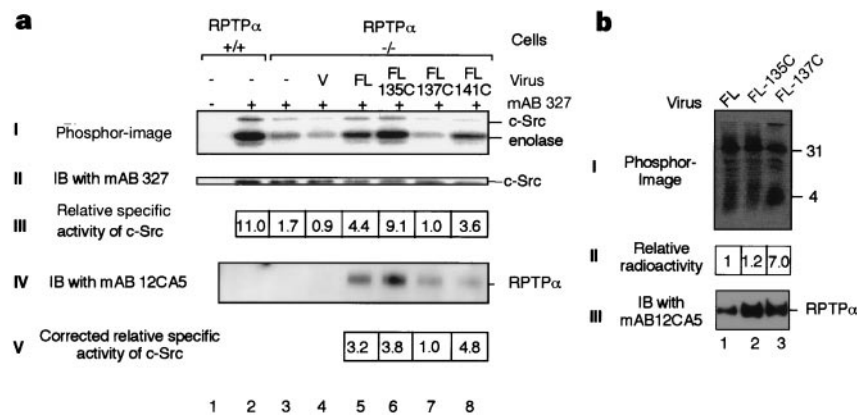


Figure 2 Mutant FL-137C has a much reduced ability to activate c-Src in RPTP α ^{-/-} cells. **a**, In vitro phosphorylation. RPTP α ^{+/+} cells and RPTP α ^{-/-} cells were either mock-infected (-), infected with control virus not encoding RPTP α (V) or infected with retrovirus expressing FL, FL-135C, FL-137C or FL-141C as indicated. Monoclonal antibody (mAb) 327 was used for c-Src immunoprecipitation from cell lysates (+) (lanes 2–8) but was omitted in the control (-) (lane 1). Panel I, Phosphorimage showing enolase phosphorylated by c-Src and autophosphorylated c-Src. Panel II, Immunoblot with mAb 327 showing the levels of c-Src in the immunoprecipitate (IPs). Panel III, The specific activity of c-Src was determined as the ratio of the level of phosphorylated enolase in panel I (quantified by phosphorimaging using ImageQuant software) to the level of c-Src in panel II (quantified by densitometry using NIH Image Version 1.57 software). Panel IV,

Immunoblot of whole cell lysates with mAb 12CA5 showing the levels of exogenously expressed HA-tagged RPTP α . Panel V, Corrected relative specific activity of c-Src was determined by normalizing the relative c-Src specificity in panel III with RPTP α expression level in panel IV (quantified by densitometry using NIH Image Version 1.57 software). Similar results were observed in three experiments, and the results of a representative experiment are shown. **b**, Analysis of c-Src phosphorylation by CNBr phosphopeptide mapping performed four days after retroviral infection as described¹². Panel I, Phosphorimage. Panel II, Quantification of the 4,000 M_r phosphorylated-Tyr529-containing phosphopeptide using ImageQuant software. Panel III, RPTP α in the ³²P-labelled lysate was immunoprecipitated and detected by IB using mAb 12CA5.

migrated at a relative molecular mass of ~ 130 K in SDS–polyacrylamide gel electrophoresis (PAGE) under reducing conditions, as determined by immunoblotting (Fig. 1b, lanes 1–5). In contrast, all of the Cys-mutant proteins, but not the wild-type protein, migrated predominantly at ~ 260 K under nonreducing conditions (Fig. 1b, lanes 6–10). This indicates that the RPTP α proteins were correctly expressed and fully glycosylated, and that the Cys mutants efficiently oligomerized *in vivo* through disulphide-bond formation. We confirmed that the RPTP α oligomers were actually homodimers because FL-135C and a truncated FL-135C mutant lacking the cytoplasmic domain formed heterodimers when coexpressed in 293 cells (data not shown).

We have been unable to assay the PTP activity of the Cys dimers *in vitro*, because the iodoacetamide included in the lysis buffer to prevent cleavage of the RPTP α dimers following cell lysis alkylates the active-site Cys and abrogates catalytic activity. In the absence of iodoacetamide, the proportion of dimeric RPTP α was significantly reduced because endogenous cellular glutathione reduces the disulphide bond and thus dissociates the dimer (data not shown). For this reason, we resorted to an *in vivo* assay for RPTP α activity, in which we determined the ability of retrovirally transduced wild-type and mutant RPTP α to dephosphorylate Tyr 529 of c-Src and thereby increase the activity of c-Src in RPTP α ^{−/−} fibroblasts derived from RPTP α knock-out mice¹⁴. RPTP α ^{−/−} cells have greatly reduced c-Src activity¹⁴. Four days after retroviral infection, when expression of RPTP α protein peaked (data not shown), infected RPTP α ^{−/−} cells were lysed (for technical reasons we have been unable to generate an FL-139C retrovirus). Endogenous c-Src protein was isolated by

immunoprecipitation using anti-Src monoclonal antibody 327, and its activity was measured by *in vitro* phosphorylation experiments using enolase as a substrate. c-Src activity was clearly detectable in RPTP α ^{+/+} cells (Fig. 2a, I and II, lane 2), being ~ 40 fold higher than that in a control without monoclonal antibody 327 (compare lanes 1 and 2, quantification not shown). As shown previously¹⁴, the specific activity of c-Src was ~ 6 fold lower in RPTP α ^{−/−} cells than in RPTP α ^{+/+} cells (Fig. 2a, I, II and III, lanes 2 and 3). Expression of the wild-type RPTP α (FL) significantly increased the specific activity of c-Src (lanes 4 and 5), as did the expression of FL-135C and FL-141C (Fig. 2a, lanes 4, 6 and 8). However, expression of FL-137C did not activate c-Src (lanes 4 and 7). Immunoblotting analysis showed that FL-137C was expressed to at least the same level as FL-141C (Fig. 2a, IV, lanes 7 and 8). In fact, FL-137C failed to activate c-Src, as well as the wild type, even when it was expressed at a higher level than the wild-type protein (see Fig. 4). When corrected for RPTP α expression levels, FL, FL-135C, and FL-141C proteins all activated c-Src to a similar extent, whereas FL-137C always activated c-Src more weakly (Fig. 2a, V, lanes 5, 6, 7 and 8). CNBr peptide mapping of *in vivo* ³²P-labelled c-Src indicates that c-Src Tyr 529 phosphorylation was higher in cells expressing FL-137C than in cells expressing FL or FL-135C, consistent with the lower Src activity in the FL-137C-expressing cells (Fig. 2b). Similar results were obtained from multiple experiments.

Immunofluorescence confocal microscopy showed indistinguishable plasma membrane staining for RPTP α in cells expressing either FL (Fig. 3b) or FL-137C (Fig. 3c), whereas plasma membrane staining was not detected in the cells infected with control retrovirus (Fig. 3a). Cell-surface labelling by biotinylation followed by immunoprecipitation of RPTP α showed that FL and FL-137C were localized to the cell surface in proportion to their expression levels (Fig. 3d, lanes 1 and 2). We therefore conclude that FL-137C was correctly localized to the plasma membrane. Together, these results indicate that dimerization through Cys at position 137, but not at position 135 or position 141, reduces the ability of RPTP α to dephosphorylate Try 529 in c-Src. Our results also indicate that an inhibited dimer may only form when the monomers in the dimer are oriented in a particular geometry with respect to one another. In other words, to be functionally inhibited, the PTP dimer may require a specific rotational coupling, as is the case for activation of Neu/ErbB2 by Cys-induced dimerization¹⁷. By chemical cross-linking, we have obtained evidence that full-length RPTP α homodimerizes extensively on the cell surface (G.J., J.d.H. and T.H., unpublished observations). However, RPTP α exists on the cell surface as populations of monomers and dimers in equilibrium,

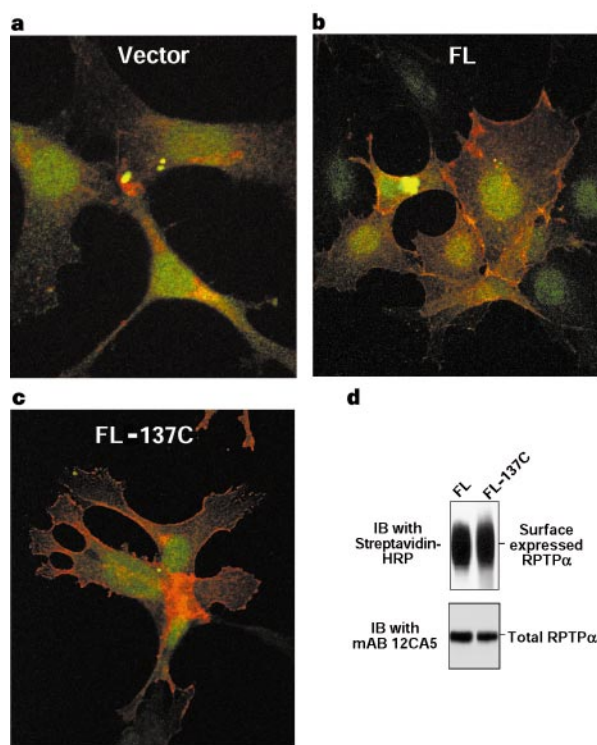


Figure 3 Membrane localization of RPTP α proteins. Immunofluorescence staining of RPTP α ^{−/−} cells infected with control retrovirus (a), retrovirus expressing FL (b) or FL-137C (c) was performed as described²³ using mAb 12CA5 and goat anti-mouse Ig-Texas Red (1:200 dilution; Southern Biotechnology, Inc.). Confocal microscopy images are shown. Red fluorescence, HA-tagged RPTP α . Green fluorescence, GFP. d, Cell-surface biotinylation of intact transiently transfected 293 cells was performed as described²⁸ using EZ-link-Sulfo-NHS-LC-Biotin (Pierce). FL and FL-137C proteins were immunoprecipitated with antiserum 5478 and detected by IB with either Streptavidin-HRP for surface-expressed RPTP α (top) or mAb 12CA5 for total RPTP α (bottom).

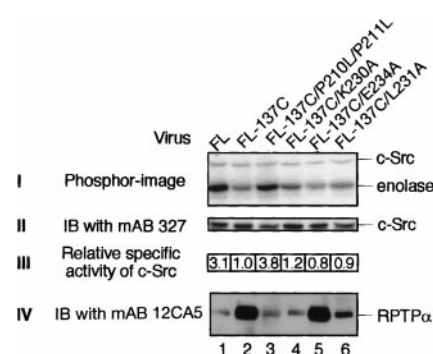


Figure 4 A P210/P211L double mutation in the wedge region activates RPTP α FL-137C. RPTP α ^{−/−} cells were infected with retrovirus expressing either FL, FL-137C, FL-137C/P210/P211L, FL-137C/K230A, FL-137C/E234A, or FL-137C/L231A RPTP α , and the relative specific activity of the c-Src in mAb 327 IPs was measured by *in vitro* phosphorylation using enolase as a substrate as described in Fig. 2. Similar results were observed in three experiments and the results of a representative experiment are shown.

and the inhibition by Cys 137-mediated dimerization therefore occurs by locking RPTP α into the inactive dimeric state.

To test whether the wedge-mediated interactions predicted by the RPTP α D1 crystal structure are needed for the loss of activity in the FL-137C dimer, we investigated the effect of wedge mutations on the activity of the FL-137C mutant protein. Mutations in the wedge region were made in FL-137C by substituting Pro 210/211 with Leu, and Lys 230, Glu 234 and Leu 231 individually with Ala. Pro 210 and 211 are in the elbow of the wedge and play an important role in holding the wedge rigid, and Lys 230 and Glu 234 make direct contacts with residues in the active site, whereas Leu 231, mutated as a control, does not appear to be involved in dimerization³. Expression of FL-137C/P210L/P211L led to activation of c-Src to a similar extent as FL RPTP α (Fig. 4, I, II and III, lanes 1 and 3), whereas expression of FL-137C, FL-137C/K230A, FL-137C/E234A and FL-137C/L231A did not activate c-Src (Fig. 4, I, II and III, lanes 2, 4, 5 and 6), even when expressed at much higher levels than the FL and FL-137C/P210L/P211L proteins (Fig. 4, IV, lanes 1, 2, 3, 5 and 6). Similar results were obtained in repeated experiments (the wedge mutants were typically expressed at lower levels than FL-137C; in Fig. 4, the expression of FL wild-type protein was atypically low, but this experiment was chosen because its level is equal to that of FL-137C/P210L/P211L). Consistent with the *in vitro* phosphorylation experiments, immunoblotting analysis using monoclonal antibody clone 28 (ref. 19), which recognizes the activated form of c-Src containing unphosphorylated Tyr 529, showed that c-Src Tyr 529 phosphorylation was lower in cells expressing FL or FL-137C/P210L/P211L than in cells expressing FL-137C and other wedge mutants (data not shown). Cell-surface labelling by biotinylation showed that all the wedge mutants were localized to the cell surface in proportion to their expression levels, as with FL and FL-137C (data not shown). This indicates that the P210/211L double mutations, but not the single mutations in the wedge region, significantly restored the activity of FL-137C. Consistent with the c-Src activity assay, chemical crosslinking experiments showed that P210/P211L double mutation reduced the dimerization efficiency of RPTP α compared with wild type (G.J., J.d.H. and T.H., unpublished observations).

In summary, our evidence indicates that RPTP α can dimerize *in vivo*, and that dimerization decreases the catalytic activity of full-length RPTP α *in vivo* through a wedge region interaction, as proposed³. This is also in keeping with results from the EGFR-CD45 dimer⁶. As other RPTPs have the potential to dimerize^{7,20}, dimerization may represent a physiologically relevant negative regulatory mechanism for a subset of RPTPs as proposed^{3,4}; however, there is structural evidence that other RPTPs may not be regulated by dimerization^{21,22}. The mechanism by which RPTP α dimerization might be regulated remains to be elucidated. Phosphorylation of Ser 180/204 in the juxtamembrane region by protein kinase C, which enhances activity, is one possible mechanism^{23,24}. Alternatively, a ligand that induces dimerization of RPTP α or stabilizes the monomeric state may exist. □

Methods

Expression vectors, site-directed mutagenesis and antibodies

The complementary DNA encoding a full length murine RPTP α (FL) with an HA epitope inserted between amino acids 19 and 20 has been described²⁵. Cys mutants FL-135C, FL-137C, FL-139C and FL-141C (Fig. 1a) were prepared using the Quikchange Site-Directed Mutagenesis Kit (Stratagene) with sense mutagenesis primers: 135C Top (5'-ccactcagaaa cctggcccccgcag-3'); 137C Top (5'-cagaaacctccctgtgcagatgagacc-3'); 139C Top (5'-ccccccgcatgtgagaccacattatgc-3'); and 141C Top (5'-cccgccagatgagatcccaattatgcg-3'). Wedge mutants in the P137C background were also prepared by mutagenesis using the following sense primers: P210/211L (5'-aggaagtactactactcctgtg-3'); D227A (5'-agaatgctgctgacaaatga-3'); K230A (5'-gatgacatgcgattctcaga-3'); E234A (5'-ctcttcagagcagaattcaac-3'); and L231A (5'-gacaataaggccttcagagaa-3'). The cDNAs were cloned into both the mammalian expression vector pSG5 (ref. 25) and the retroviral expression vector pCR2. pCR2 is a bicistronic vector based on the pCL series²⁶ (C.A. Joazeiro, unpublished data), and the recombinant vector expresses green fluorescent protein (GFP) and RPTP α simultaneously. 12CA5 is a mouse monoclonal antibody against the HA epitope; 5478 is a

rabbit polyclonal antiserum raised against a glutathione-S-transferase (GST)-fusion protein containing the entire cytoplasmic domain of murine RPTP α (ref. 23); 327 is a mouse monoclonal antibody raised against v-Src that also recognizes c-Src²⁷.

Cells, transient transfections, retrovirus packaging and infections

RPTP $\alpha^{-/-}$ cells and RPTP $\alpha^{+/+}$ cells are embryonic fibroblasts derived from RPTP α knock-out mice and normal (control) mice, respectively¹⁴. These and HEK 293 cells were cultured in Dulbecco's-modified Eagle's medium supplemented with 10% fetal bovine serum at 37 °C and 10% CO₂. Transient transfection of 293 cells was done by calcium phosphate precipitation¹². For protein expression, cells were harvested 48 h after transfection. For retrovirus production, 15 μ g recombinant retrovirus vector and 10 μ g ecotropic packaging vector pCLECO²⁶ were co-transfected into 293 cells. Media containing retrovirus were harvested at 36, 48 and 60 h by replenishing 3 ml of fresh medium after each harvest, and the virus-containing media were pooled and filtered. For infection, RPTP $\alpha^{-/-}$ cells were seeded onto 50-mm tissue-culture dishes 24 h before infection with the virus-containing media containing 5 μ g ml⁻¹ polybrene every 12 h, a total of three times. Generally, nearly 100% of the cells were GFP positive and infected, as judged by fluorescence-activated cell sorting analysis (data not shown).

Immunoprecipitation and *in vitro* phosphorylation

RPTP $\alpha^{-/-}$ cells were seeded onto 50-mm plates and infected with retrovirus as described above. RPTP $\alpha^{+/+}$ cells, when used, were exposed to polybrene but not to retrovirus. Cells were lysed in 0.75 ml RIPA buffer²³. The levels of exogenously overexpressed RPTP α were determined by immunoblotting whole cell lysates using monoclonal antibody 12CA5. The c-Src was immunoprecipitated with anti-Src monoclonal antibody 327 and its activity was assayed by *in vitro* phosphorylation experiments using enolase as substrate as described¹². The levels of c-Src in the immunoprecipitates were determined by immunoblotting using monoclonal antibody 327.

Received 5 July; accepted 26 August 1999.

- Hunter, T. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* **80**, 225–236 (1995).
- Tonks, N. K. & Neel, B. G. From form to function: signaling by protein tyrosine phosphatases. *Cell* **87**, 365–368 (1996).
- Bilwes, A. M., den Hertog, J., Hunter, T. & Noel, J. P. Structural basis for inhibition of receptor protein-tyrosine phosphatase α by dimerization. *Nature* **382**, 555–559 (1996).
- Weiss, A. & Schlessinger, J. Switching signals on or off by receptor dimerization. *Cell* **94**, 277–280 (1998).
- Desai, D. M., Sap, J., Schlessinger, J. & Weiss, A. Ligand-mediated negative regulation of a chimeric transmembrane receptor tyrosine phosphatase. *Cell* **73**, 541–554 (1993).
- Majeti, R., Bilwes, A. M., Noel, J. P., Hunter, T. & Weiss, A. Dimerization-induced inhibition of receptor protein tyrosine phosphatase function through an inhibitory wedge. *Science* **279**, 88–91 (1998).
- Felberg, J. & Johnson, P. Characterization of recombinant CD45 cytoplasmic domain proteins. Evidence for intramolecular and intermolecular interactions. *J. Biol. Chem.* **273**, 17839–17845 (1998).
- Wallace, M. J., Fladd, C., Batt, J. & Rotin, D. The second catalytic domain of protein tyrosine phosphatase δ (PTP δ) binds to and inhibits the first catalytic domain of PTP α . *Mol. Cell. Biol.* **18**, 2608–2616 (1998).
- Sap, J., D'Eustachio, P., Givol, D. & Schlessinger, J. Cloning and expression of a widely expressed receptor tyrosine phosphatase. *Proc. Natl Acad. Sci. USA* **87**, 6112–6116 (1990).
- Wu, L., Buist, A., den Hertog, J. & Zhang, Z. Y. Comparative kinetic analysis and substrate specificity of the tandem catalytic domains of the receptor-like protein-tyrosine phosphatase α . *J. Biol. Chem.* **272**, 6994–7002 (1997).
- Harder, K. W., Moller, N. P., Peacock, J. W. & Jirik, F. R. Protein-tyrosine phosphatase α regulates Src family kinases and alters cell-substratum adhesion. *J. Biol. Chem.* **273**, 31890–31900 (1998).
- den Hertog, J. et al. Receptor protein tyrosine phosphatase α activates pp60^{c-src} and is involved in neuronal differentiation. *EMBO J.* **12**, 3789–3798 (1993).
- Zheng, X. M., Wang, Y. & Pallen, C. J. Cell transformation and activation of pp60^{c-src} by overexpression of a protein tyrosine phosphatase. *Nature* **359**, 336–339 (1992).
- Su, J., Muranjan, M. & Sap, J. Receptor protein tyrosine phosphatase α activates Src-family kinases and controls integrin-mediated responses in fibroblasts. *Curr. Biol.* **9**, 505–511 (1999).
- Ponniah, S., Wang, D. Z., Lim, K. L. & Pallen, C. J. Targeted disruption of the tyrosine phosphatase PTP α leads to constitutive downregulation of the kinases Src and Fyn. *Curr. Biol.* **9**, 535–538 (1999).
- Cao, H., Bangalore, L., Dompe, C., Bormann, B. J. & Stern, D. F. An extra cysteine proximal to the transmembrane domain induces differential cross-linking of p185^{neu} and p185^{neu'}. *J. Biol. Chem.* **267**, 20489–20492 (1992).
- Burke, C. L. & Stern, D. F. Activation of Neu (ErbB-2) mediated by disulfide bond-induced dimerization reveals a receptor tyrosine kinase dimer interface. *Mol. Cell. Biol.* **18**, 5371–5379 (1998).
- Sorokin, A., Lemmon, M. A., Ullrich, A. & Schlessinger, J. Stabilization of an active dimeric form of the epidermal growth factor receptor by introduction of an inter-receptor disulfide bond. *J. Biol. Chem.* **269**, 9752–9759 (1994).
- Kawakatsu, H. et al. A new monoclonal antibody which selectively recognizes the active form of Src tyrosine kinase. *J. Biol. Chem.* **271**, 5680–5685 (1996).
- Takeda, A., Wu, J. J. & Maizel, A. L. Evidence for monomeric and dimeric forms of CD45 associated with a 30-kDa phosphorylated protein. *J. Biol. Chem.* **267**, 16651–16659 (1992).
- Hoffmann, K. M., Tonks, N. K. & Barford, D. The crystal structure of domain 1 of receptor protein-tyrosine phosphatase μ . *J. Biol. Chem.* **272**, 27505–27508 (1997).
- Nam, H. J., Poy, F., Krueger, N. X., Saito, H. & Frederick, C. A. Crystal structure of the tandem phosphatase domains of RPTP LAR. *Cell* **97**, 449–457 (1999).
- Tracy, S., van der Geer, P. & Hunter, T. The receptor-like protein-tyrosine phosphatase, RPTP α , is phosphorylated by protein kinase C on two serines close to the inner face of the plasma membrane. *J. Biol. Chem.* **270**, 10587–10594 (1995).

24. den Hertog, J., Sap, J., Pals, C. E., Schlessinger, J. & Kruijer, W. Stimulation of receptor protein-tyrosine phosphatase α activity and phosphorylation by phorbol ester. *Cell. Growth Differ.* **6**, 303–307 (1995).
25. den Hertog, J. & Hunter, T. Tight association of GRB2 with receptor protein-tyrosine phosphatase α is mediated by the SH2 and C-terminal SH3 domains. *EMBO J.* **15**, 3016–3027 (1996).
26. Naviaux, R. K., Costanzi, E., Haas, M. & Verma, I. M. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J. Virol.* **70**, 5701–5705 (1996).
27. Lipsich, L. A., Lewis, A. J. & Brugge, J. S. Isolation of monoclonal antibodies that recognize the transforming proteins of avian sarcoma viruses. *J. Virol.* **48**, 352–360 (1983).
28. Levy-Toledano, R., Caro, L. H., Hindman, N. & Taylor, S. I. Streptavidin blotting: a sensitive technique to study cell surface proteins; application to investigate autophosphorylation and endocytosis of biotin-labeled insulin receptors. *Endocrinology* **133**, 1803–1808 (1993).

Acknowledgements

We are grateful to C. Joazeiro for providing the pCR2 retroviral vector and to W. Eckhart and P. Blume-Jensen for suggestions and critical review of the manuscript. This work is supported by grants from the National Cancer Institute (T.H.) and Dutch Cancer Society (J.d.H.). T.H. is a Frank and Else Schilling American Cancer Society Research Professor. G.J. is supported by a fellowship from the American Cancer Society.

Correspondence and requests for materials should be addressed to G.J. (e-mail: gjiang@aim.salk.edu).

Phytochrome signalling is mediated through nucleoside diphosphate kinase 2

Giltsu Choi*, Hankuil Yi*, Jaeho Lee†, Yong-Kook Kwon‡, Moon Soo Soh*, Byongchul Shin*, Zigmund Luka*, Tae-Ryong Hahn† & Pill-Soon Song*‡§

* Kumho Life and Environmental Science Laboratory, Kwangju 500-712, Korea

† Department of Genetic Engineering, Kyung Hee University, Suwon 449-701, Korea

‡ Department of Chemistry, University of Nebraska, Lincoln, Nebraska 68588, USA

§ Department of Life Science, Kwangju Institute of Science and Technology, Kwangju 500-712, Korea

Because plants are sessile, they have developed intricate strategies to adapt to changing environmental variables, including light. Their growth and development, from germination to flowering, is critically influenced by light, particularly at red (660 nm) and far-red (730 nm) wavelengths^{1,2}. Higher plants perceive red and far-red light by means of specific light sensors called phytochromes (A–E)³. However, very little is known about how light signals are transduced to elicit responses in plants. Here we report that nucleoside diphosphate kinase 2 (NDPK2) is an upstream component in the phytochrome signalling pathway in the plant *Arabidopsis thaliana*. In animal and human cells, NDPK acts as a tumour suppressor⁴. We show that recombinant NDPK2 in *Arabidopsis* preferentially binds to the red-light-activated form of phytochrome *in vitro* and that this interaction increases the activity of recombinant NDPK2. Furthermore, a mutant lacking NDPK2 showed a partial defect in responses to both red and far-red light, including cotyledon opening and greening. These results indicate that NDPK2 is a positive signalling component of the phytochrome-mediated light-signal-transduction pathway in *Arabidopsis*.

Biochemical approaches have identified the heterotrimeric guanine-nucleotide-binding protein Ca^{2+} /CaM and cyclic GMP as components of phytochrome-mediated light-signal-transduction pathways³. Indeed, the phosphorylation and dephosphorylation of

specific proteins, including phytochrome, may be involved^{5,6}. Genetic approaches have identified ten pleiotropic mutants (*cop/fus/det*) that display constitutively photomorphogenic phenotypes and several other photomorphogenic mutants⁷; some mutants specific to a certain phytochrome signal-transduction pathway have also been identified^{8–13}. However, the recently identified PIF3 and PKS1 are the only signalling components that interact directly with phytochrome to have been thoroughly investigated^{14,15}.

We used a yeast two-hybrid screening method to identify proteins associated with phytochrome A. As the bait, we used the carboxy-terminal domain of *Arabidopsis* phytochrome A (position 568–1193). From the screening, we identified 52 yeast colonies that expressed both reporter genes (*his3⁺/lacZ⁺*) in a phytochrome-A-dependent manner, three of which encoded phytochrome A (two clones, position 588–1193; one clone, position 943–1193) (data not shown). Identification of the expected phytochrome A clones from the screening indicated that the C-terminal domain of phytochrome A has the correct folding structure in yeast, so it can be used to identify other phytochrome-associated proteins.

Sequence analysis indicated that 26 phytochrome-A-associated proteins are NDPK2. NDPK2 interacted specifically with phytochrome A, but not with DNA-binding domain of GAL4 alone or those fused to other proteins such as laminin C and p53 (Fig. 1a).

We established whether two spectral forms of phytochrome interact differently with NDPK2 by using purified recombinant *Arabidopsis* NDPK2 and oat phytochrome (Fig. 1b). Recombinant

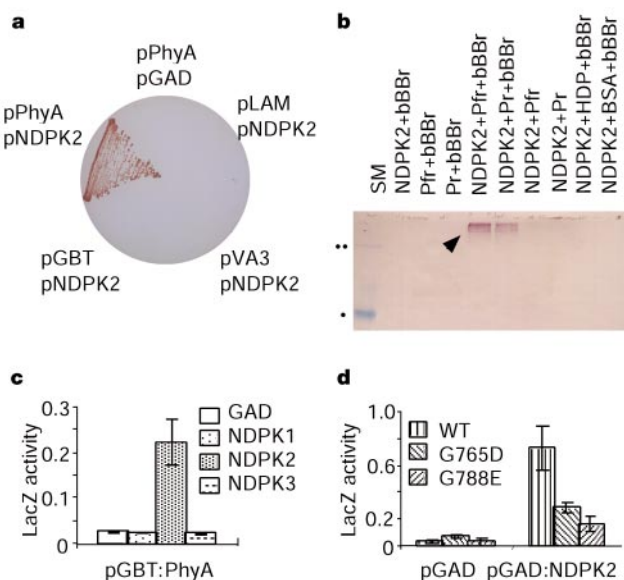


Figure 1 NDPK2 interacts preferentially with the P_F form of phytochrome, whereas NDPK1 and NDPK3 do not interact. **a**, Specific interaction between NDPK2 and phytochrome A. pPhyA, position 568–1193 of *Arabidopsis* phytochrome A in pGBT9; pLAM and pVA3, laminin and p53 in pGBT9; pNDPK2, the full-length NDPK2 in pGAD424; pGBT, pGBT9. Yeasts were plated on HIS-dropout plates. **b**, *In vitro* crosslinking assay between NDPK2 and phytochrome. Arrow indicates crosslinked products of high relative molecular mass; 250 K (two dots) and 98 K (one dot) size markers (SM) are indicated. NDPK2, recombinant NDPK2; bBBr, the crosslinking agent bisbromobimane; Pfr and Pr, different spectral forms (P_F and P_R) of phytochrome; HDP, heat-denatured phytochrome; BSA, bovine serum albumin. **c**, ONPG assay showing that only NDPK2 interacts with phytochrome A. Genes encoding NDPK were in pGAD424 vectors (NDPK1, NDPK2 and NDPK3). GAD, a control vector containing only the GAL4 activation domain. Position 656–1193 of wild-type phytochrome A (pGBT:PhyA) was used as bait. LacZ activity was expressed as the ratio of absorbance at 420 and 600 nm. Error bar, standard deviation. **d**, ONPG assay showing that two identified missense mutations in phytochrome A disrupt the interaction with NDPK2 (pGAD:NDPK2). Position 656–1193 of wild-type (WT) and mutant (G765D, G788E) phytochrome A in pGBT9 was used. Error bar, standard deviation.

NDPK2 interacted preferentially with the far-red phytochrome (P_{fr}) form of purified oat phytochrome, as assayed by crosslinking using bisbromobimane. Quantitatively, P_{fr} bound 78% better than the red phytochrome (P_r) form, according to the band intensity. Because NDPKs are encoded by a small gene family, we tested whether the interaction is specific to NDPK2. Unlike NDPK2 (GenBank accession no. AF017640), NDPK1 (accession no. AF017641) and NDPK3 (accession no. AF044265) did not interact with the C-terminal domain of phytochrome A (Fig. 1c). Furthermore, two identified missense mutations in phytochrome A (G765D and G788E) disrupted the ability of phytochrome A to bind to NDPK2 (Fig. 1d). Taken together, the results suggest that NDPK2 does bind to phytochrome A.

If two proteins interact, they are expected to share the same subcellular space. To investigate the subcellular localization of NDPK2, we generated transgenic tobaccos that expressed fusion proteins of NDPK2 with green fluorescent protein (GFP). We determined the localization of these NDPK2–GFP fusion proteins in transgenic tobacco plants by using fluorescence microscopy. NDPK2–GFP is localized mainly in the nucleus and cytoplasm (Fig. 2). In the cytoplasm, NDPK2–GFP is concentrated in the regions surrounding the chloroplasts. A similar localization has been reported for NDPK1a, which has the same 78 amino acids as

NDPK2 at its amino-terminal (two genes are 98.7% identical for the whole region)¹⁶. NDPK2 and phytochromes therefore do share the same subcellular space.

To investigate whether phytochrome modulates the activity of NDPK2, the γ -phosphate-exchanging activity of NDPK2 was examined spectrophotometrically by monitoring the enzyme-coupled decrease in NADH in the presence of various concentrations of the P_{fr} or P_r form of oat phytochrome. A significant increase in NDPK2 activity was observed in the presence of the P_{fr} form (Fig. 3a). The P_{fr} form of oat phytochrome (150 nM) increased the activity of NDPK2 by about 70% over activity without phytochrome, whereas the P_r form had no effect on NDPK2 activity. However, neither P_{fr} nor P_r modulated the activity of recombinant NDPK1 under the same conditions. These results are consistent with the yeast two-hybrid data that the C-terminal domain of *Arabidopsis* phytochrome A interacts specifically with NDPK2 but not with NDPK1. To investigate the biochemical mechanism for the increased activity of NDPK2 by the P_{fr} form, we determined the K_m value of NDPK2 by using deoxycytidine diphosphate (dCDP). The K_m value of NDPK2 decreased from 0.60 mM to 0.43 mM in the presence of the P_{fr} form of phytochrome (Fig. 3b), indicating that phytochrome increases NDPK2 activity by decreasing the K_m value in a red-light-dependent manner.

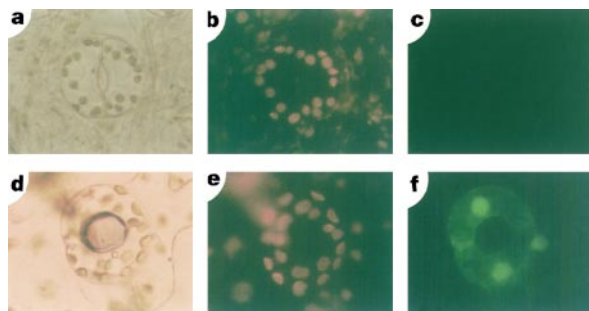


Figure 2 Subcellular localization of NDPK2. The localization of NDPK2–GFP fusion proteins was determined by using leaf epidermal peelings of transgenic tobacco lines. **a–c**, Results from a non-transgenic tobacco line; **d–f**, from a transgenic line containing the NDPK2–GFP fusion gene. **a, d**, Stomata cells on the epidermal peelings; **b, e**, the same cells showing chloroplast fluorescence; **c, f**, the same cells showing GFP fluorescence.

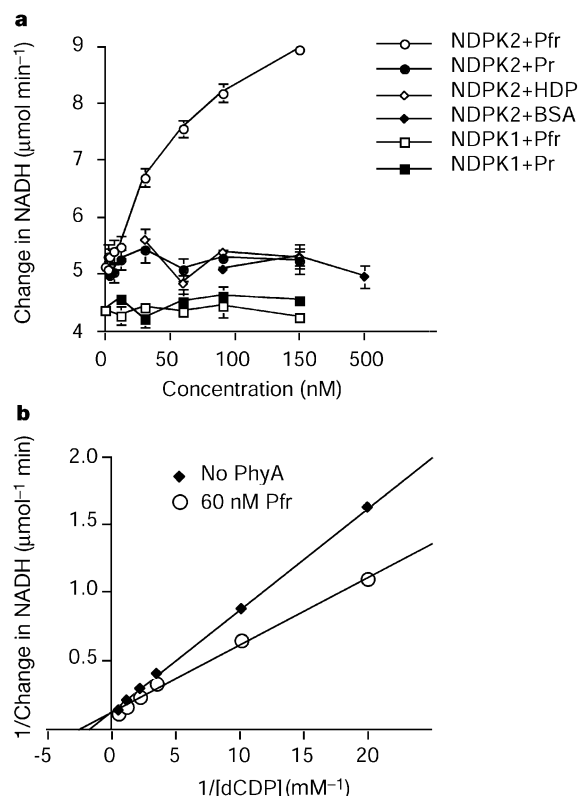


Figure 3 The P_{fr} -dependent activation of the recombinant *Arabidopsis* NDPK2. **a**, The γ -phosphate-exchange activity of NDPK2 and NDPK1 in the presence of different spectral forms of phytochrome or other control proteins. **b**, Lineweaver–Burk plot of NDPK2 activity.

To investigate the role of NDPK2 *in planta*, we isolated a T-DNA insertion mutant of NDPK2 that had T-DNA inserted in an intron region between amino acids 128 and 129 of NDPK2 (data not shown). The homozygous *ndpk2* mutant grew well, indicating that the NDPK2 gene may not be required under laboratory conditions. Because NDPK2 is a phytochrome-A-associated protein, we investigated the development of phytochrome-A-mediated seedlings under continuous far-red light. The *ndpk2* mutant had a slightly shorter hypocotyl under all conditions (Fig. 4a). When we compared the rate of decrease in hypocotyl length by increasing the fluence rate, the decrease reduced only marginally in the mutant (Fig. 4b), indicating that NDPK2 is not essential for phytochrome-A-mediated hypocotyl elongation or that its function is replaceable by others. However, the mutant showed defects in cotyledon opening and hook straightening under continuous far-red light, indicating that the *ndpk2* mutant partly phenocopies *phyA* mutants in these aspects. Wild-type seedlings developed an open hook and cotyledons at $7 \mu\text{W cm}^{-2}$, whereas *ndpk2* mutant seedlings had a closed hook and cotyledons at this fluence rate (Fig. 4c). At $14.7 \mu\text{W cm}^{-2}$, the hook was open but the cotyledons were not completely open, indicating that the mutant may not completely abolish the phytochrome-A-mediated far-red response. These results indicate that NDPK2 is involved in the phytochrome-A-mediated far-red high-irradiance response.

The C-terminal domains of phytochromes A and B are functionally interchangeable¹⁷, raising the possibility that NDPK2 might also be involved in other phytochrome-mediated light-signal transductions. The *ndpk2* mutant was marginally less sensitive for red light in terms of hypocotyl elongation, but the mutant showed a strong defect in red-light-induced cotyledon opening and greening (Fig. 5), indicating that NDPK2 is involved in the phytochrome-B-mediated red-light high-irradiance response. Unlike PIF3, however, a mutation in NDPK2 primarily affected cotyledon opening and greening. Another phytochrome-binding protein, PKS1, has been proposed to inhibit phytochrome B signalling. These findings suggest that phytochromes interact with various effector molecules and regulate their functions to modulate a range of responses to light.

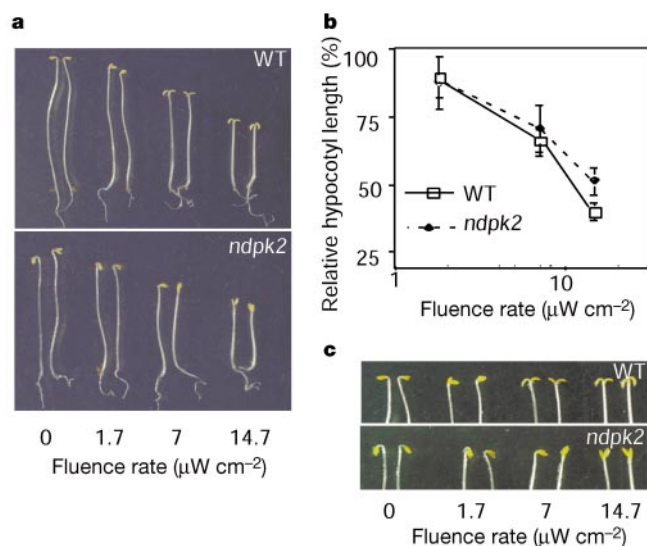


Figure 4 Far-red high-irradiance response of a *ndpk2* mutant. **a**, Seedlings of a wild type and a *ndpk2* mutant grown under continuous far-red light for 5 days. **b**, Percentage decrease of hypocotyl length in wild type and a *ndpk2* mutant by increasing far-red-light fluence rate. **c**, Seedlings of wild type and a *ndpk2* mutant showing hook and cotyledon development under far-red light.

NDPK is required for many eukaryotic developmental processes. In *Drosophila*, for example, mutation of the NDPK homologues called *awd* (for *abnormal wing disc*) results in abnormal cell morphology, aberrant differentiation and cell necrosis^{18,19}. In mammals, NDPK, which is also referred as Nm23, acts as a tumour suppressor⁴. Several enzyme activities, including as a histidineaspartic acid kinase, a serine/threonine kinase, a transcription factor and a G-protein activator, have been proposed to explain its role in development^{20–23}. However, little is known about the role of this multifunctional NDPK in plant development. NDPK1a, which is very similar to NDPK2, has recently been shown to complement the *gcn4* mutant of yeast and binds to the HIS4 promoter to activate transcription¹⁶. Red-light-dependent phosphorylation of NDPK in peas also has been reported²⁴. Our data show that the P_{fr} form of purified oat phytochrome increases the activity of recombinant NDPK2 and that NDPK2 is required for phytochrome signalling *in planta*. The molecular mechanism of modulation by NDPK2 of phytochrome-mediated cotyledon opening remains to be further investigated. □

Methods

Yeast two-hybrid screening

An *Arabidopsis thaliana* cDNA library for the yeast two-hybrid system was constructed from messenger RNAs purified from 1-week-old, light-grown seedlings (Col-0 ecotype). The library was constructed in pGAD424 vector (Clontech). As a bait, position 568–1193 of *Arabidopsis* phytochrome A (a gift from P. H. Quail) was subcloned into pGBT9 vector (Clontech). Transformants (1.5×10^6) were screened and plasmids isolated from the positive clones were re-introduced into yeast together with the bait or control vectors (pGBT9, pLAM 5' and pVA3; Clontech) to check the interaction and its specificity. To investigate whether some of mutations identified in phytochrome A disrupt the interaction, position 656–1193 of two phytochrome mutants (G765D and G788E) was subcloned into pGBT9 vector and the interaction was measured by o-nitrophenyl- β -D-galactopyranoside (ONPG) assay, for which three independent colonies were picked and measured for enzyme activities.

Enzymatic assays

Oat phytochrome was purified from a etiolated oat seedlings as described²⁵. *Arabidopsis* NDPK1 (amino acids 1–115) and NDPK2 (79–226) were subcloned into pET28a vector (Novagen) and the recombinant proteins purified from *Escherichia coli* over Ni²⁺-agarose column according to manufacture's guidelines (Qiagen). NDPK2 with 78 amino acids

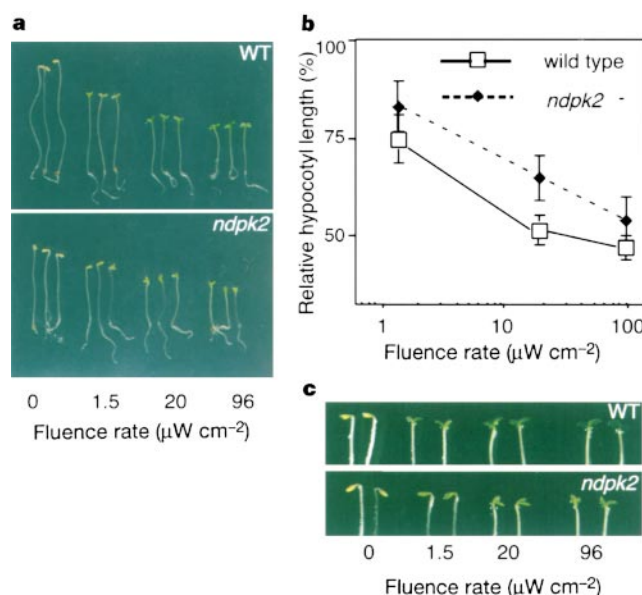


Figure 5 Red-light high-irradiance response of a *ndpk2* mutant. **a**, Seedlings of wild type and a *ndpk2* mutant grown under continuous red light for 5 days. **b**, Percentage decrease of hypocotyl length in wild type and a *ndpk2* mutant by increasing red-light fluence rate. **c**, Seedlings of wild type and a *ndpk2* mutant showing cotyledon opening and greening under red light.

deleted retains the ability to bind to phytochrome both in yeast and *in vitro* (data not shown). The γ -phosphate-exchanging activity of NDPK was measured as described²⁶ with minor modifications. The assay buffer contained 100 mM Tris-Cl, pH 7.5, 100 mM KCl, 25 mM MgCl₂, 3 mM phosphoenolpyruvate, 2 mM ATP, 0.3 mM NADH, 2.5 units pyruvate kinase, 2.5 units lactate dehydrogenase and 1 mM dCDP. The reaction was started by adding purified NDPK to a final concentration of 3 nM in 600 μ l at 20 °C. NDPK activity was measured by monitoring the lactate dehydrogenase-pyruvate kinase-coupled NADH decrease (change in absorption at 340 nm) in the reaction mixture. The effect of phytochrome on NDPK activity was examined by incubating purified oat phytochromes, illuminated by red (660 nm, P_{fr} form) or far-red (730 nm, P_r form) light, with NDPK in the reaction mixture. The K_m value of NDPK2 with dCDP was measured in a similar manner except that the final concentration of phytochromes was fixed at 60 nM.

In vitro binding assay

Six-histidine-tagged NDPK2 (500 ng) was crosslinked with the P_r or P_{fr} form of oat phytochrome (1 μ g) by adding the thiol-reactive crosslinking reagent bisbromobimane (bBBBr) to a final concentration of 1 mM in 20 μ l reaction buffer (100 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM MgCl₂)²⁷. After 1 min of crosslinking, the reaction was stopped by adding glutathione to a final concentration 100 mM. The crosslinked samples were separated on SDS-PAGE and detected using a histidine-tag antibody.

Localization of NDPK2

A NDPK2 gene was fused to GFP and subcloned into pBI121 plant binary vector (Clontech). Transgenic tobacco plants were then generated by *Agrobacterium*-mediated transformation and the localization of GFP-fused NDPK2 was observed by a fluorescence microscope (Olympus).

Characterization of a *ndpk2* mutant

A *ndpk2* mutant was isolated from Feldmann T-DNA insertion lines²⁸ and an insertion site identified by sequencing. The homozygous line of the *ndpk2* mutant was used for the characterization. To investigate the far-red high-irradiance response, both wild-type (WS-2) and *ndpk2* mutants were grown on MS agar plates without sucrose under continuous far-red light (8 μ W cm⁻²) for 5 days.

Received 30 April; accepted 3 August 1999.

1. von Arnim, A. & Deng, X.-W. Light control of seedling development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 215–243 (1996).
2. Fankhauser, C. & Chory, J. Light control of plant development. *Annu. Rev. Cell Dev. Biol.* **13**, 203–229 (1997).
3. Bowler, C., Neuhaus, G., Yamagata, H. & Chua, N. H. Cyclic GMP and calcium mediate phytochrome phototransduction. *Cell* **77**, 73–81 (1994); erratum, *ibid* **79**, 743 (1994).
4. de la Rosa, A., Williams, R. L. & Steeg, P. S. Nm23/nucleotide diphosphate kinase: toward a structural and biochemical understanding of its biological functions. *Bioessays* **17**, 53–62 (1995).
5. Ellich, T. D. & Chory, J. Photochrome: if it looks and smells like a histidine kinase, is it a histidine kinase? *Cell* **91**, 713–716 (1997).
6. Yeh, K. C., Wu, S. H., Murphy, J. T. & Lagarias, J. C. A cyanobacterial phytochrome two-component light sensory system. *Science* **277**, 1505–1508 (1997).
7. Kwok, S. F., Piekos, B., Misera, S. & Deng, X. W. A complement of ten essential and pleiotropic *Arabidopsis* COP/DET/FUS genes is necessary for repression of photomorphogenesis in darkness. *Plant Physiol.* **110**, 731–742 (1996).
8. Whitelam, G. C. *et al.* Phytochrome A null mutants of *Arabidopsis* display a wild-type phenotype in white light. *Plant Cell* **5**, 757–768 (1993).
9. Ahmad, M. & Cashmore, A. R. The *pef* mutants of *Arabidopsis thaliana* define lesions early in the phytochrome signaling pathway. *Plant J.* **10**, 1103–1110 (1996).
10. Barnes, S. A., Quaggio, R. B., Whitelam, G. C. & Chua, N. H. *hy1* defines a branch point in phytochrome a signal transduction pathways for gene expression. *Plant J.* **10**, 1155–1161 (1996).
11. Wagner, D., Hoecker, U. & Quail, P. H. RED1 is necessary for phytochrome B-mediated red light-specific signal transduction in *Arabidopsis*. *Plant Cell* **9**, 731–743 (1997).
12. Hoecker, U., Xu, Y. & Quail, P. H. SPA1: a new genetic locus involved in phytochrome A-specific signal transduction. *Plant Cell* **10**, 19–33 (1998).
13. Soh, M. S., Hong, S. H., Hanzawa, H., Furuya, M. & Nam, H. G. Genetic identification of FIN2, a far red light-specific signaling component of *Arabidopsis thaliana*. *Plant J.* **16**, 411–419 (1998).
14. Ni, M., Tepperman, J. M. & Quail, P. H. PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell* **95**, 657–667 (1998).
15. Fankhauser, C. *et al.* PKS1, a substrate phosphorylated by phytochrome that modulates light signaling in *Arabidopsis*. *Science* **284**, 1539–1541 (1999).
16. Zimmermann, S. *et al.* UV responsive genes of *Arabidopsis* revealed by similarity to the Gcn4 mediated UV response in yeast. *J. Biol. Chem.* **274**, 17017–17024 (1999).
17. Wagner, D., Fairchild, C. D., Kuhn, R. M. & Quail, P. H. Chromophore-bearing NH2-terminal domains of phytochromes A and B determine their photosensory specificity and differential light lability. *Proc. Natl Acad. Sci. USA* **93**, 4011–4015 (1996).
18. Dearolf, C. R., Hersperger, E. & Shearn, A. Developmental consequence of *awd3*, a cell autonomous lethal mutation of *Drosophila* induced by hybrid dysgenesis. *Dev. Biol.* **13**, 203–229 (1988).
19. Biggs, J., Hersperger, E., Steeg, P. S., Liotta, L. A. & Shearn, A. A *Drosophila* gene that is homologous to a mammalian gene associated with tumor metastasis codes for a nucleoside diphosphate kinase. *Cell* **63**, 933–940 (1990).
20. Engel, M. *et al.* A novel serine/threonine-specific protein phosphotransferase activity of Nm23/nucleotide-diphosphate kinase. *Eur. J. Biochem.* **234**, 200–207 (1995).
21. Lu, Q., Park, H., Egger, L. A. & Inouye, M. Nucleoside-diphosphate kinase-mediated signal transduction via histidyl- aspartyl phosphorelay systems in *Escherichia coli*. *J. Biol. Chem.* **271**, 32886–32893 (1996).

22. Wagner, P. D., Steeg, P. S. & Vu, N. D. Two-component kinase-like activity of nm23 correlates with its motility-suppressing activity. *Proc. Natl Acad. Sci. USA* **94**, 9000–9005 (1997).
23. Bominaar, A. A., Molijn, A. C., Pestel, M., Veron, M. & Van Haastert, P. J. Activation of G-proteins by receptor-stimulated nucleoside diphosphate kinase in *Dictyostelium*. *EMBO J.* **12**, 2275–2279 (1993).
24. Ogura, T., Tanaka, N., Yabe, N., Komatsu, S. & Hasunuma, K. Characterization of protein complexes containing nucleoside diphosphate kinase with characteristics of light signal transduction through phytochrome in etiolated pea seedlings. *Photochem. Photobiol.* **69**, 397–403 (1999).
25. Lapko, V. N. & Song, P. S. A simple and improved method of isolation and purification for native oat phytochrome. *Photochem. Photobiol.* **62**, 194–198 (1995).
26. Sommer, D. & Song, P. S. A plant nucleoside diphosphate kinase homologous to the human Nm23 gene product: purification and characterization. *Biochim. Biophys. Acta* **1222**, 464–470 (1994).
27. Kosower, E. M. & Kosower, N. S. Bromobimane probes for thiols. *Methods Enzymol.* **251**, 133–148 (1995).
28. Feldmann, K. A. T-DNA insertion mutagenesis in *Arabidopsis*-mutational spectrum. *Plant J.* **1**, 71–82 (1991).

Acknowledgements

This work was supported by the Kumho Petrochemical Co., the Korea Science and Engineering Foundation, The Academic Research Fund of the Ministry of Education, Republic of Korea and USPHS-NIH. We thank the *Arabidopsis* Stock Centre for sending seeds.

Correspondence and requests for materials should be addressed to G.C. (gtcoi@ksc.kumho.co.kr) or P.S. (pssong@ksc.kumho.co.kr).

Reconstitution of actin-based motility of *Listeria* and *Shigella* using pure proteins

Thomas P. Loisel, Rajaa Boujemaa, Dominique Pantaloni & Marie-France Carlier

Dynamique du Cytosquelette, LEBS, CNRS, Gif-sur-Yvette, France

Actin polymerization is essential for cell locomotion and is thought to generate the force responsible for cellular protrusions. The Arp2/3 complex is required to stimulate actin assembly at the leading edge in response to signalling^{1–6}. The bacteria *Listeria* and *Shigella* bypass the signalling pathway and harness the Arp2/3 complex to induce actin assembly and to propel themselves in living cells^{7–10}. However, the Arp2/3 complex alone is insufficient to promote movement. Here we have used pure components of the actin cytoskeleton to reconstitute sustained movement in *Listeria* and *Shigella in vitro*. Actin-based propulsion is driven by the free energy released by ATP hydrolysis linked to actin polymerization, and does not require myosin. In addition to actin and activated Arp2/3 complex, actin depolymerizing factor (ADF, or cofilin) and capping protein are also required for motility as they maintain a high steady-state level of G-actin, which controls the rate of unidirectional growth of actin filaments at the surface of the bacterium. The movement is more effective when profilin, α -actinin and VASP (for *Listeria*) are also included. These results have implications for our understanding of the mechanism of actin-based motility in cells.

Attempts to understand how site-directed actin polymerization can produce movement in cells have centred on work to connect actin to the signalling pathway in *Listeria* and *Shigella*. The Arp2/3 complex is activated by all proteins of the WASP/Scar family^{1–5}, which connect the cytoskeleton to the signalling pathway^{3,11}. The *Shigella* IcsA protein recruits N-WASP¹² and activates it in a Cdc42-like fashion, leading to Arp2/3 activation¹⁰. The *Listeria* ActA protein activates Arp2/3 by mimicking N-WASP⁹. N-WASP can mediate the insertional polymerization of actin because its amino-terminal domain binds F-actin and its carboxy-terminal domain

binds G-actin in a profilin-like fashion, shuttling actin subunits to the growing barbed ends¹⁰. The combination of ActA and VASP at the surface of *Listeria* may fulfil a function equivalent to that of IcsA-activated N-WASP¹³. However, when bacteria are brought into a solution of pure F-actin, assembled at steady state in the presence of ATP, Arp2/3 and either VASP or N-WASP, the bacteria initiate actin assembly at their surface but fail to move^{8,10,13}, whereas they can move steadily in cell-free extracts with an ATP supply. We sought to reconstitute actin-based movement by supplementing the bacterium-activated machinery with proteins that would shuttle actin subunits to the barbed ends generated by locally activated Arp2/3 complex.

We observed sustained movement of *Listeria* and N-WASP-coated *Escherichia coli* (IcsA) (Fig. 1) in a physiological ionic strength buffer solution, pH 7.5, containing F-actin (7.5 μM) supplemented with 0.1 μM Arp2/3, 0.05 μM capping protein, 5 μM ADF, 2.5 μM profilin and 0.25 μM α -actinin, and the *Listeria* motility medium also contained 0.5 μM VASP; these concentrations are generally compatible with those found *in vivo*. The bacteria formed actin tails about 30 μm long and moved at rates of up to 2 $\mu\text{m min}^{-1}$ (*E. coli* (IcsA)) or 4 $\mu\text{m min}^{-1}$ (*Listeria*), similar to values recorded in extracts, and the percentage of motile bacteria was also the same as in extracts¹³.

This motility medium was developed because actin-based movement results from the rapid turnover of actin filaments, which grow from their barbed ends at specific sites and depolymerize from the pointed ends of all filaments in the medium in a treadmilling process¹⁴. Efficient propulsion first implies that the treadmilling is strongly biased, or 'funnelled'¹⁵, such that barbed ends grow only where force is needed. Complete barbed-end capping of all fila-

ments in the medium fulfils this condition and maintains G-actin subunits at the high concentration required (the critical concentration) at the pointed end. Furthermore, barbed ends generated at the bacterial surface must be capped as soon as they dissociate from the motile ActA-VASP or IcsA-N-WASP machinery. In addition, the treadmilling rate is governed by the rate of depolymerization from filament pointed ends. This reaction is very slow in pure F-actin solutions, which is why bacteria fail to move in standard F-actin solution. Actin depolymerizing factor and profilin are therefore expected to contribute to actin-based propulsion¹⁶⁻¹⁹, as they act synergistically, leading to a 125-fold increase in filament turnover and to a large increase in the steady-state amount of ATP-G-actin and profilin-actin. The F-actin solution containing capping protein, ADF and profilin therefore acts as a chemostat to elicit the rapid shuttling of ATP-G-actin subunits, supporting actin-based propulsion of *Listeria* as well as *Shigella*. The high steady-state concentration of ATP-G-actin represents the 'fuel' for different bacteria-associated motors. The F-actin present in the medium buffers the ATP-G-actin at its steady-state concentration, so it may be present at any concentration. Bacterial movement also requires α -actinin²⁰, and the role of VASP in *Listeria* movement has been studied^{13,21,22}. The non-pathogenic *E. coli* expressing IcsA was used as a valid substitute for *Shigella*, as it can move in *Xenopus* egg extracts^{23,24} and in N-WASP-supplemented platelet extracts¹⁰.

We established the optimum composition of the dynamic steady-state F-actin solution suitable for *Listeria* or *E. coli* (IcsA) movement and determined which proteins are essential for motility by measuring the rates of bacterial propulsion when a range of concentrations of Arp2/3, capping protein, ADF, profilin, α -actinin (and VASP for *Listeria*) were added to an F-actin solution (Table 1). In addition to

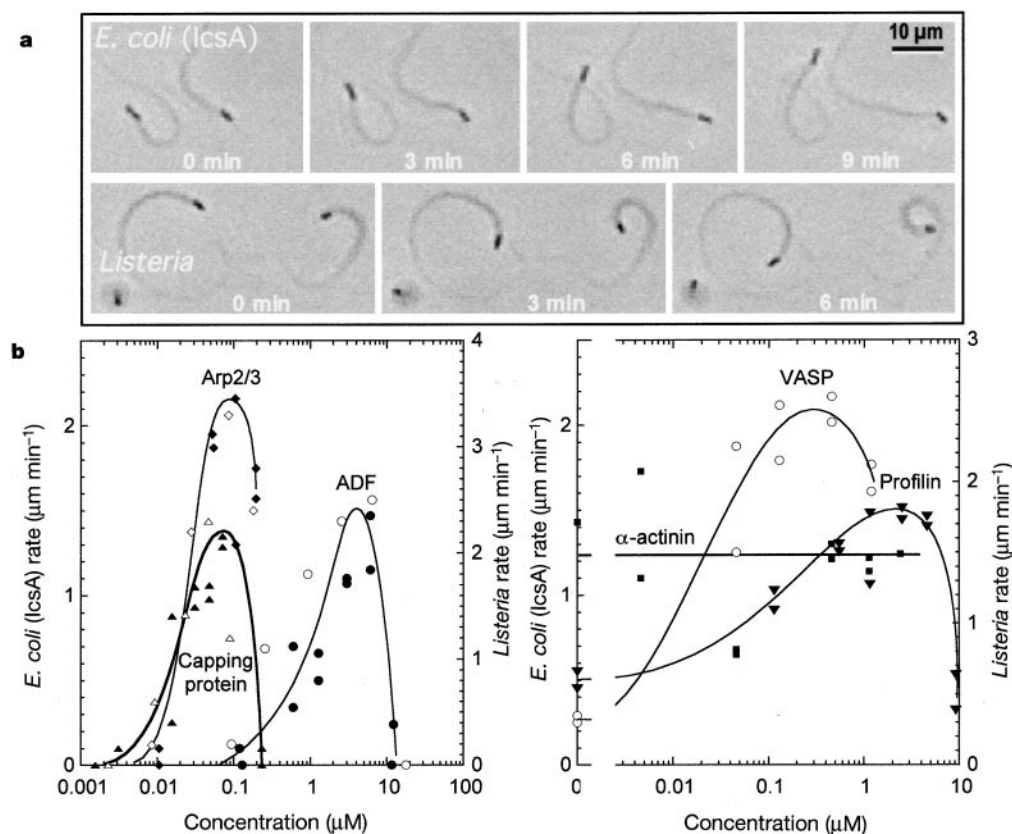


Figure 1 Reconstitution of actin-based movement of *E. coli* (IcsA) and *L. monocytogenes* with pure components. **a**, Time-lapse phase-contrast video-microscopy recording of bacteria movement. *E. coli* (IcsA) moved at average rates of 1 $\mu\text{m min}^{-1}$ and *Listeria* at average rates of 2.2 $\mu\text{m min}^{-1}$. **b**, Dependence of bacterial rate on the concentration of

essential components (left) and non-essential components (right). Filled symbols, *E. coli* (IcsA); open symbols, *Listeria*. Note the logarithmic abscissa scale. Duplicate data points represent independent measurements made by two experimentalists. Each data point is the average of 10 measurements. The standard deviation was generally 20%.

Table 1 Optimum concentrations of proteins involved in the motility assay

Bacteria	Protein	Optimal concentration range	Proteins concentration (μM)							Maximum rate ($\mu\text{m min}^{-1}$)
			Actin (μM)	Arp2/3	ADF	Capping protein	Profilin	α -Actinin	VASP	
<i>E. coli</i> IcsA	Arp2/3	(0.05–0.1)	7.3		5.1	0.07	2.5	0.25	0	2.2
	ADF	(2–6)	7.3	0.08		0.03	1.1	0.47	0	1.5
	CP	(0.03–0.1)	7.3	0.08	5.0		1.1	0.44	0	1.4
	Profilin	(0.5–5)	7.3	0.1	5.1	0.07		0.46	0	1.5
	α -Actinin	(0.2–3)	7.3	0.1	5.1	0.07	2.5		0	1.3
	–	–	7.8	0.09	2.7	0.06	0	0	0	1.2
<i>Listeria</i>	Arp2/3	(0.05–0.1)	7.4		2.6	0.05	2.5	0.24	0.42	3.3
	ADF	(2–6)	7.4	0.065		0.05	2.5	0.24	0.42	2.5
	CP	(0.03–0.1)	7.4	0.065	2.6		2.5	0.24	0.42	2.3
	VASP	(0.1–1)	7.4	0.09	5.1	0.07	2.2	0.24		2.6
	–	–	7.4	0.085	2.6	0.05	0	0	0	0.5
	–	–								

actin, Arp2/3 and N-WASP (for *E. coli* (IcsA)), ADF and capping protein were essential for movement. In the presence of just Arp2/3, ADF and capping protein, *Listeria* and N-WASP-coated *E. coli* (IcsA) moved slowly at rates of 0.5 and 1.2 $\mu\text{m min}^{-1}$, respectively.

The rate of movement varied with all protein concentrations with a bell-shaped curve. Free Arp2/3 present in the motility medium is inactive^{2–5,10}, being transiently activated at the bacterium surface by interaction with ActA⁹ or IcsA-bound N-WASP¹⁰ to nucleate/branch actin filaments in a spatially restricted fashion^{6,9} and then incorporated at branches in the actin meshwork¹⁰. The dependence of the rate on ADF reached a maximum at half-saturation of F-actin by ADF, like the treadmilling rate *in vitro*¹⁷ and consistent with biochemical data showing that ADF increases the turnover of actin filaments even when Arp2/3 is present¹⁹. The requirement of barbed-end capping for movement points to a function for capping proteins that explains¹⁶ their reported effect in cell motility^{25–27}. At very high concentrations, however, capping protein blocks the elongation of actin filaments formed at the bacterium surface, overcoming the barbed-end nucleation process elicited by Arp2/3, so movement is slower. *Listeria* and *E. coli* (IcsA) depended equally on ADF, profilin and capping protein for movement, although *Listeria* moved twice as fast as *E. coli* (IcsA) under all conditions, owing to their different motile machineries at the surface (ActA–VASP and IcsA–N-WASP, respectively). Profilin (and VASP for *Listeria*), although not absolutely essential, were nonetheless important for movement. Profilin increases the rate of movement two- to threefold, consistent with previous motility assays^{13,28} and with biochemical data showing that profilin enhances the processivity of treadmilling in the presence of ADF¹⁸. The

absence of VASP was more deleterious than that of profilin in *Listeria* movement, causing a tenfold decrease in the rate of movement, even though normal actin tails were still formed. This result is in agreement with a recent report showing that VASP is present at the leading edge in amounts correlating with the rate of protrusion²⁹, but is only in partial agreement with genetic and depletion studies^{13,21,22} showing that abnormal actin tails form and movement ceases for most bacteria in the absence of VASP. In cells and cell extracts, other proteins might interact with ActA at the VASP-binding site, either compensating for or aggravating the effects of VASP removal. Our reconstituted motility medium therefore represents a better tool for testing the specific functions of different proteins in motility. We found that α -actinin was not required for movement, and varying its concentration had no effect on the rate of movement. However, when α -actinin was not present, the actin tails were much less dense and the bacteria attached to their tails drifted appreciably in the solution, presumably owing to the lack of connections between the filaments in the medium and those that form the actin tails. Vinculin was not required for *E. coli* (IcsA) movement, in agreement with a previous report³⁰.

In conclusion, actin-based movement can be reconstituted *in vitro* by using a small number of actin-binding proteins whose function in the regulation of actin dynamics is well established. No myosin motor is required for actin-based movement. Actin-based movement operates by using the free energy released by ATP hydrolysis to actin assembly. Knowing which proteins are essential for actin-based movement is an important step toward understanding its detailed mechanism and its regulation by the signalling pathway. □

Methods

Proteins

Actin was purified from rabbit muscle¹⁸, profilin from bovine spleen¹⁸, capping protein from bovine erythrocytes³¹ and Arp2/3 complex from bovine brain¹⁰. Human ADF was bacterially expressed and purified¹⁷. Human histidine-tagged VASP¹³ and N-WASP¹⁰ were expressed in Sf21 cells, and α -actinin was from Sigma. An SDS–PAGE electrophoresis pattern of the proteins is shown in Fig. 2.

Bacteria

L. monocytogenes Lut12 pactA3^{13,28} and *E. coli* (IcsA)¹⁰ were grown and used in motility assays.

Motility assays

The motility medium contained: 10 mM Tris–Cl, pH 7.5, 1.5 mM Mg–ATP, 5 mM DTT, 0.1 M KCl, 7.5 μM F-actin, 0.1 μM Arp2/3 complex, 0.05 μM capping protein, 5 μM ADF, 2.5 μM profilin, 0.5 μM α -actinin, 0.5 μM VASP (for *Listeria* only), 0.35% methylcellulose, 5 mg ml^{−1} BSA and 2×10^8 bacteria ml^{−1}. *E. coli* (IcsA) were preincubated with 1 μM N-WASP, centrifuged and resuspended¹⁰ before being placed in the motility medium. Samples for phase-contrast optical-microscopy observation of movement were prepared as described^{10,13}. Rates of movement were measured by determining the coordinates of the centre of gravity of the bacterium (1 pixel resolution) every 5 s.

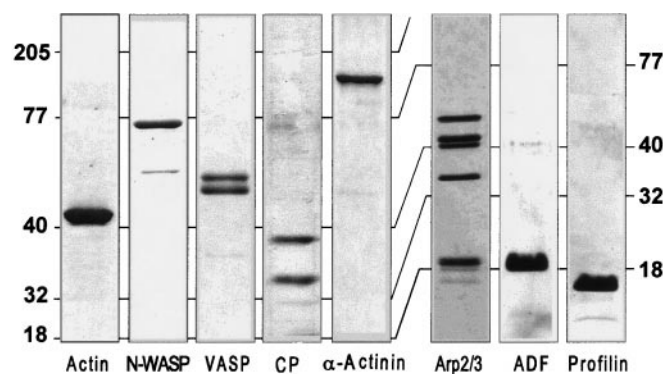


Figure 2 SDS–PAGE pattern of proteins used in the reconstitution assay of actin-based motility. The left five lanes were 10% and the right three lanes 15% acrylamide gels. CP, capping protein.

Received 4 June; accepted 28 July 1999.

- Machesky, L. M. & Insall, R. H. Scar1 and the related Wiskott-Aldrich syndrome protein WASP regulate the actin cytoskeleton through the Arp2/3 complex. *Curr. Biol.* **8**, 1347–1356 (1998).
- Machesky, L. M. *et al.* Scar, a WASP-related protein, activates dendritic nucleation of actin filaments by the Arp2/3 complex. *Proc. Natl Acad. Sci. USA* **96**, 3739–3744 (1999).
- Rohatgi, R. *et al.* The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell* **97**, 221–231 (1999).
- Winter, D., Lechler, T. & Li, R. Activation of the Arp2/3 complex by Bee 1p, a WASP-family protein. *Curr. Biol.* **9**, 501–504 (1999).
- Yarar, D., To, W., Abo, A. & Welch, M. D. The Wiskott-Aldrich syndrome protein directs actin-based motility by stimulating actin nucleation with the Arp2/3 complex. *Curr. Biol.* **9**, 555–558 (1999).
- Svitkina, T. M. & Borisy, G. G. Arp2/3 complex and Actin Depolymerizing Factor/cofilin in dendritic organization and treadmilling of actin filament array in lamellipodia. *J. Cell Biol.* **145**, 1009–1026 (1999).
- Theriot, J. A., Mitchison, T. J., Tilney, L. G. & Portnoy, D. A. The rate of actin-based motility of intracellular *Listeria monocytogenes* equals the rate of actin polymerisation. *Nature* **357**, 257–260 (1992).
- Welch, M. D., Iwamatsu, A. & Mitchison, T. J. Actin polymerization is induced by Arp2/3 protein complex at the surface of *Listeria monocytogenes*. *Nature* **385**, 265–269 (1997).
- Welch, M. D., Rosenblatt, J., Skoble, J., Portnoy, D. A. & Mitchison, T. J. Interaction of human Arp2/3 complex and the *Listeria monocytogenes* ActA protein in actin filament nucleation. *Science* **281**, 105–108 (1998).
- Egile, C. *et al.* Activation of the Cdc42 effector N-WASP by the *Shigella* IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. *J. Cell Biol.* **146**, 1319–1332 (1999).
- Miki, H., Sasaki, T., Takai, Y. & Takenawa, T. Induction of filopodium formation by a N-WASP-related actin-depolymerizing factor. *Nature* **391**, 93–96 (1998).
- Suzuki, T., Miki, H., Takenawa, T. & Sasakawa, C. Neural Wiskott-Aldrich syndrome protein is implicated in the actin-based motility of *Shigella flexneri*. *EMBO J.* **17**, 2767–2776 (1998).
- Laurent, V. *et al.* Role of the proteins of the Ena/VASP family in actin-based motility of *Listeria monocytogenes*. *J. Cell Biol.* **144**, 1245–1258 (1999).
- Small, J. V. Getting the actin filaments straight: nucleation-release or treadmilling. *Trends Cell Biol.* **5**, 52–55 (1995).
- Carlier, M.-F. & Pantaloni, D. Control of actin dynamics in cell motility. *J. Mol. Biol.* **269**, 459–467 (1997).
- Carlier, M.-F. Control of actin dynamics. *Curr. Opin. Cell Biol.* **10**, 45–51 (1998).
- Carlier, M.-F. *et al.* Actin Depolymerising Factor (ADF/cofilin) enhances the rate of filament turnover: implication in actin-based motility. *J. Cell Biol.* **136**, 1307–1322 (1997).
- Didry, D., Carlier, M.-F. & Pantaloni, D. Synergy between ADF and profilin in enhancing actin filament turnover. *J. Biol. Chem.* **273**, 25602–25611 (1998).
- Ressad, F., Didry, D., Egile, C., Pantaloni, D. & Carlier, M.-F. Control of actin filament length and turnover by actin depolymerizing factor (ADF/cofilin) in the presence of capping proteins and Arp2/3 complex. *J. Biol. Chem.* **274**, 20970–20976 (1999).
- Nanavati, D., Ashton, F. T., Sanger, J. M. & Sanger, J. W. Dynamics of actin and α -actinin in the tails of *Listeria monocytogenes* in infected PTK₂ cells. *Cell Motil. Cytoskel.* **28**, 346–358 (1994).
- Smith, G. A., Theriot, J. A. & Portnoy, D. A. The tandem repeat domain in the *Listeria monocytogenes* ActA protein controls the rate of actin-based motility, the percentage of moving bacteria and the localization of VASP and profilin. *J. Cell Biol.* **35**, 647–660 (1996).
- Lasa, I. *et al.* Identification of two regions in the N-terminal domain of ActA involved in the actin tail formation by *Listeria monocytogenes*. *EMBO J.* **16**, 1531–1540 (1997).
- Goldberg, M. & Theriot, J. A. *Shigella flexneri* surface protein IcsA is sufficient to direct actin-based motility. *Proc. Natl Acad. Sci. USA* **92**, 6572–6576 (1995).
- Kocks, C. *et al.* The unrelated surface proteins ActA of *Listeria monocytogenes* and IcsA of *Shigella flexneri* are sufficient to confer actin-based motility on *Listeria innocua* and *E. coli* respectively. *Mol. Microbiol.* **18**, 413–423 (1995).
- Cunningham, C., Stosel, T. P. & Kwiatkowski, D. Enhanced motility in NIH 3T3 fibroblasts that overexpress gelsolin. *Science* **251**, 1233–1236 (1991).
- Hug, C. *et al.* Capping protein levels influence actin assembly and cell motility in *Dictyostelium*. *Cell* **81**, 591–600 (1995).
- Sun, H., Kwiatkowski, K., Wooten, D. & Yin, H. Effects of CapG overexpression on agonist-induced motility and second messenger generation. *J. Cell Biol.* **129**, 147–156 (1995).
- Marchand, J.-B. *et al.* Actin-based movement of *Listeria monocytogenes*: actin assembly results from the local maintenance of uncapped filament barbed ends at the bacterium surface. *J. Cell Biol.* **130**, 331–343 (1995).
- Rottner, K., Behrendt, B., Small, J. V. & Wehland, J. VASP dynamics during lamellipodia protrusion. *Nature Cell Biol.* **1**, 321–322 (1999).
- Goldberg, M. *Shigella* actin-based motility in the absence of vinculin. *Cell Motil. Cytoskel.* **37**, 44–53 (1997).
- Kuhlmann, P. A. & Fowler, V. M. Purification and characterization of an α 1 β 2 isoform of CapZ from human erythrocytes: cytosolic location and inability to bind to Mg⁺⁺ ghosts suggest that erythrocyte actin filaments are capped by adducin. *Biochemistry* **36**, 13461–13472 (1997).

Acknowledgements

We thank V. Laurent and C. Egile for discussion, D. Didry for Arp2/3 complex purification, and F. Ressad for human ADF. We acknowledge partial support from the Association pour la Recherche contre le Cancer, the Association Française contre les Myopathies and a Human Frontier in Science grant. T.P.L. is supported by a fellowship from the Natural Sciences and Engineering Research Council of Canada.

Correspondence and requests for materials should be addressed to M.-F.C. (e-mail: carlier@lebs.caps-gif.fr).

14-3-3 σ is required to prevent mitotic catastrophe after DNA damage

Timothy A. Chan, Heiko Hermeking, Christoph Lengauer, Kenneth W. Kinzler & Bert Vogelstein

The Johns Hopkins Oncology Center, Program in Human Genetics, and The Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, 424 N. Bond Street, Baltimore, Maryland 21231, USA

14-3-3 σ is a member of a family of proteins that regulate cellular activity by binding and sequestering phosphorylated proteins. It has been suggested that 14-3-3 σ promotes pre-mitotic cell-cycle arrest following DNA damage, and that its expression can be controlled by the p53 tumour suppressor gene¹. Here we describe an improved approach to the generation of human somatic-cell knockouts, which we have used to generate human colorectal cancer cells in which both 14-3-3 σ alleles are inactivated. After DNA damage, these cells initially arrested in the G2 phase of the cell cycle, but, unlike cells containing 14-3-3 σ , the 14-3-3 σ ^{-/-} cells were unable to maintain cell-cycle arrest. The 14-3-3 σ ^{-/-} cells died ('mitotic catastrophe') as they entered mitosis. This process was associated with a failure of the 14-3-3 σ -deficient cells to sequester the proteins (cyclin B1 and cdc2) that initiate mitosis and prevent them from entering the nucleus. These results may indicate a mechanism for maintaining the G2 checkpoint and preventing mitotic death.

Many cell types treated with DNA-damaging agents arrest in the G2 phase of the cell cycle. The 14-3-3 σ gene is transcriptionally activated by p53 after DNA damage, and exogenous overexpression of 14-3-3 σ can block cells in G2, indicating that 14-3-3 σ might be a component of the G2 checkpoint¹. To test this hypothesis, we created cell lines deficient in 14-3-3 σ by homologous recombination. The human 14-3-3 σ genomic locus was cloned and fully sequenced, and subclones were used to construct a targeting vector containing a geneticin-resistance gene in place of genomic 14-3-3 σ sequences. Two loxP sites were introduced into the construct, one immediately 5' and the other directly 3' to the geneticin-resistance gene. A primer site for polymerase chain reaction (PCR) derived from deleted genomic sequences was placed between the 5' loxP site and the 5' end of the selection marker². The two loxP sites and the primer-containing construct enabled a relatively rapid and robust approach to screening clones for correctly targeted alleles, and for sequentially generating clones with both alleles disrupted (see Methods).

The human colorectal cancer cell (CRC) line HCT116 was used for these experiments because it expresses wild-type p53 and 14-3-3 σ , has intact DNA-damage checkpoints and is suitable for targeted homologous recombination³. Two independently derived 14-3-3 σ ^{+/-} and three 14-3-3 σ ^{-/-} clones were used for the studies described below (Fig. 1a–c). All clones of the same genotype behaved identically.

We investigated the function of 14-3-3 σ in human cells following DNA damage by adriamycin, a commonly used chemotherapeutic agent, as well as by ionizing radiation. The disruption of 14-3-3 σ had a marked effect on the responses of the cells to these agents. After either treatment, 14-3-3 σ ^{+/-} and 14-3-3 σ ^{-/-} cells increased in size, with a corresponding nuclear enlargement and DNA content characteristic of cells arrested in G2 (Fig. 2a). Although 14-3-3 σ ^{-/-} cells appeared to enter into a similar G2 block following DNA damage, they failed to maintain this arrest and eventually underwent 'mitotic catastrophe' (Fig. 2a). Mitotic catastrophe is an apoptosis-like process that begins in prophase, after dissolution of the nuclear membrane, and is associated with the entry of cdc2 and cyclin B1 into the nucleus^{4,5}. This process began ~24 h after

treatment and was complete after 72–96 h (Fig. 3a). Although chromatin condensation and micronucleation occurred during mitotic catastrophe, this process was distinguished from more common forms of apoptosis by a relative lack of DNA degradation, indicated by both DNA laddering (not shown) and TdT-mediated dUTP nick end labelling (TUNEL) staining for nicked DNA (compare with ceramide-induced apoptosis in Fig. 2b).

Roughly 75% of parental cells and 14-3-3 $\sigma^{+/-}$ heterozygote cells arrested after DNA damage with a G2 DNA content, whereas the remaining 25% of cells arrested in G1 (Fig. 3b). A similar fraction (~25%) of 14-3-3 $\sigma^{-/-}$ cells arrested in G1, but a relative decrease in the G2 fraction was evident as early as 24 h after adriamycin treatment (Fig. 3b). At later times, the 14-3-3 $\sigma^{-/-}$ cells showed a marked decrease in cells arrested in G2 and a corresponding increase in the number of sub-G1 cells, whereas the fraction of G2-arrested cells remained constant in both 14-3-3 $\sigma^{+/+}$ and 14-3-3 $\sigma^{+/-}$ cells

(Fig. 3b). The number of G1-arrested cells remained constant in cells of all three genotypes, indicating that the cells were probably undergoing catastrophe directly from the G2/M phase rather than from all parts of the cell cycle.

To study the expression of 14-3-3 σ following DNA damage, we generated an antibody specific for 14-3-3 σ using an amino-acid sequence found in 14-3-3 σ but not in the other six 14-3-3 family members. The specificity of this antibody was verified by the detection of a protein of the expected size upon western blotting of parental cells treated with DNA-damaging agents but not of 14-3-3 $\sigma^{-/-}$ cells (Fig. 1d). The 14-3-3 σ protein was induced as early as 6 h after exposure to adriamycin, with kinetics similar to those of the p53-induced gene *p21* in the same cells (Fig. 3c).

The 14-3-3 σ antibody was used to study the protein *in situ*. There was no detectable staining in most exponentially growing 14-3-3 $\sigma^{+/+}$ cells, although a small population of large cells stained

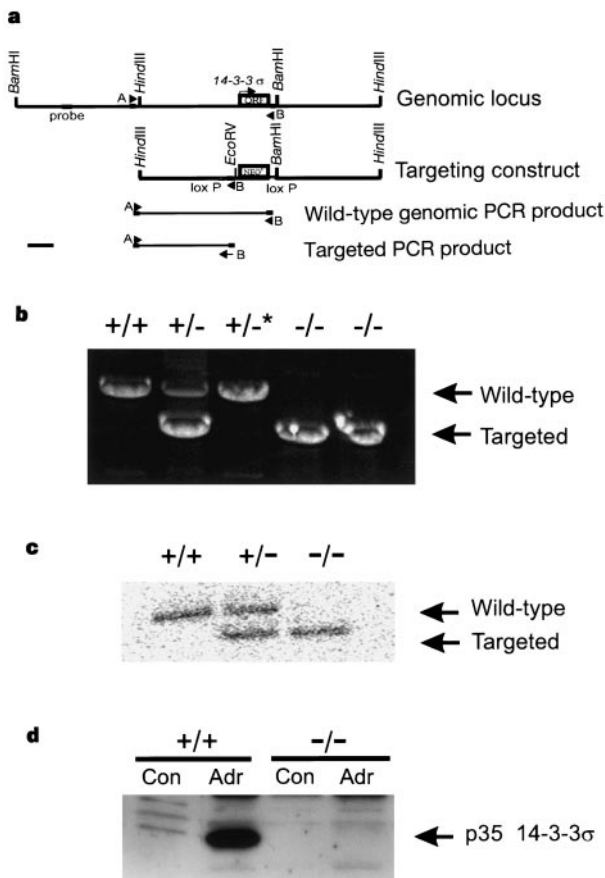


Figure 1 Generation of human cells deficient in 14-3-3 σ . **a**, Genomic structure of the human 14-3-3 σ gene and targeting construct. The box designated 'probe' represents the region used for Southern blotting. Primers at sites A and B were used in the PCR screen to identify targeted clones. The neomycin-resistance gene within the targeting construct is flanked by two loxP sites. After expression of *cre* recombinase to excise the selection cassette, the same targeting construct was used to disrupt the second 14-3-3 σ allele. Primer site B and an *Eco*RV site were incorporated directly upstream of the neomycin-resistance gene. The PCR products expected from wild-type and targeted alleles are shown at the bottom. Scale bar, 1 kb. **b**, PCR analysis of DNA from targeted cells of the indicated genotypes, using primers A and B. The 14-3-3 $\sigma^{+/-*}$ clone was derived from a 14-3-3 $\sigma^{+/-}$ clone that was infected with a recombinant adenovirus encoding the *cre* recombinase to remove the neomycin-resistance gene and primer B sequences from the targeted allele. **c**, Southern blot analysis of genomic DNA from cells with the indicated 14-3-3 σ genotype. Genomic DNA was digested with *Bam*HI and *Eco*RV and probed with the labelled 200-bp fragment shown in **a**. **d**, Western blot analysis. Cells with the indicated 14-3-3 σ genotype were cultured in the absence (Con) or presence (Adr) of adriamycin and subsequently lysed in sample buffer. Extracts were fractionated by SDS-PAGE and probed with an antibody specific for 14-3-3 σ .

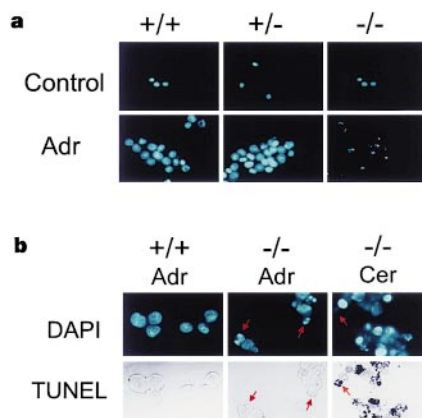


Figure 2 Mitotic catastrophe in 14-3-3 $\sigma^{-/-}$ cells. **a**, Nuclear morphology after adriamycin treatment for 72 h and staining with the DNA-specific dye DAPI. 14-3-3 $\sigma^{+/+}$ and $^{+/-}$ cells arrested and became enlarged when treated with adriamycin. 14-3-3 $\sigma^{-/-}$ cells developed condensed, fragmented chromatin when treated with the drug. Similar results were obtained after γ -irradiation (data not shown). **b**, TUNEL staining. 14-3-3 $\sigma^{+/+}$ cells treated with adriamycin underwent cell-cycle arrest but their chromatin remained uncondensed, and they did not stain with TUNEL. 14-3-3 $\sigma^{-/-}$ cells treated with adriamycin developed condensed chromatin (red arrows) but did not stain with TUNEL. Cells treated with C2-ceramide underwent micronucleation but did stain with TUNEL. The peroxidase reaction products of TUNEL staining significantly quenched DAPI fluorescence (orange arrow).

intensely (Fig. 3d). The size and morphology of the cells expressing 14-3-3 σ in the absence of DNA damage suggested that they were arrested in G2 and perhaps senescent. When cells were treated with γ -ionizing radiation or adriamycin, all the cells expressed 14-3-3 σ (Fig. 3d). As expected, 14-3-3 $\sigma^{-/-}$ cells did not stain for 14-3-3 σ either before or after treatment with adriamycin (Fig. 3d), confirming the specificity of the antibodies for immunohistochemical analysis.

A key step in regulating the progression of cells from G2 into mitosis is the activation of the protein kinase cdc2 (ref. 6). Cdc2 and cyclin B1 are cytoplasmic during interphase and are translocated to the nucleus in late G2 to initiate mitosis^{4,7}. The DNA damage checkpoint prevents the activation of cdc2–cyclin B1 complexes and thereby prevents cells from entering a potentially lethal mitosis when their DNA is damaged⁸. When we treated parental cells or 14-3-3 $\sigma^{+/-}$ cells with adriamycin or ionizing radiation, cdc2 and cyclin B1 remained in the cytoplasm throughout the duration of the experiment (Fig. 4a). In identically treated 14-3-3 σ -deficient cells, cdc2 and cyclin B1 were initially cytoplasmic, as in the parental cells; however, cdc2 and cyclin B1 soon migrated to the nucleus in most of the cells, and mitotic catastrophes occurred soon after (Fig. 4a, b). Normally, the entry of cyclin B1, cdc2, and cdc25C into the nucleus occurs simultaneously⁴; however, cdc25C remained localized in the cytoplasm before nuclear-membrane dissolution, despite the fact that cyclin B1 and cdc2 were in the nucleus (Fig. 4a). Double-staining with antibodies specific for cdc2 and 14-3-3 σ in DNA-damaged cells revealed a striking colocalization of the two proteins in the cytoplasm of parental cells (Fig. 4c). This may indicate that 14-3-3 σ normally sequesters cyclin B1 and cdc2 in the

cytoplasm, keeping cdc2–cyclin B1 from entering the nucleus and initiating mitosis following DNA damage. To investigate the molecular determinants of this sequestration, we carried out immunoprecipitation experiments using anti-14-3-3 σ antibodies. Western blots of the immunoprecipitates showed that cdc2 and cyclin B1 were bound to 14-3-3 σ , and that this binding markedly increased following adriamycin treatment (Fig. 4d; and data not shown). Wee1 kinase was also found in the immunoprecipitates (Fig. 4d); it has already been shown that cdc2 and cyclin B1 can bind to each other, and that wee1 binds to cdc2–cyclin B1 (ref. 9). No cdc25C could be detected (Fig. 4d), indicating that other 14-3-3 family members, but not 14-3-3 σ , may be responsible for the reported cdc25C interactions¹⁰.

Thus, it appears that 14-3-3 σ normally sequesters cdc2–cyclin B1 complexes in the cytoplasm during G2 arrest, and that the absence of 14-3-3 σ eventually allows cdc2–cyclin B1 complexes to enter the nucleus, resulting in mitotic catastrophe. This idea is consistent with several previous experiments which show that the cytoplasmic localization of cdc2–cyclin B1 is a critical regulator of the G2/M transition⁸, and that mitotic catastrophes occur if mitosis is experimentally induced in the presence of unreplicated or damaged DNA⁴. Our results also support the idea that the initiation of G2 arrest is distinct from its maintenance¹¹. Notably, we observed identical mitotic catastrophes in DNA-damaged parental cells after treating them with leptomycin B, an agent that inhibits nuclear export and results in a similar lack of cytoplasmic sequestration of cdc2–cyclin B1 (data not shown). These results, in combination with previous studies, suggest a simple but redundant set of mechanisms for preventing lethal mitosis following exposure of cells to DNA-

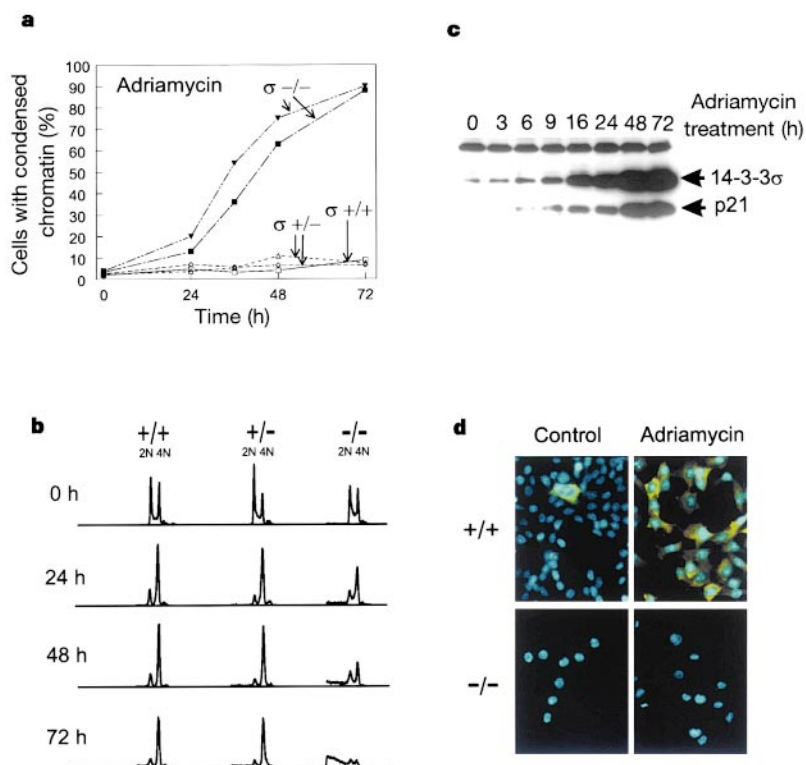


Figure 3 Cells deficient in 14-3-3 σ are sensitive to DNA-damage. **a**, Cells with the indicated 14-3-3 σ genotypes were treated with adriamycin for various time periods, stained with DAPI and scored for chromosome condensation. Two independent clones of homozygous (14-3-3 $\sigma^{-/-}$) and heterozygous (14-3-3 $\sigma^{+/-}$) cells were tested. **b**, Cells were adriamycin treated for the indicated times, stained with Hoechst 33258 and analysed by flow cytometry. The peaks labelled 2N and 4N represent cells with DNA contents characteristic of those in the G1 and G2 phases of cell cycle, respectively. **c**,

Kinetics of 14-3-3 σ induction. HCT116 cells were treated with adriamycin for the indicated time periods, and the extracts were used for western blot analysis, using antibodies specific for 14-3-3 σ and p21. **d**, Immunostaining for 14-3-3 σ . Cells of the indicated genotypes were incubated in the presence (adriamycin) or absence (control) of adriamycin, stained with a fluorescein-labelled anti-14-3-3 σ antibody (green) and counterstained with DAPI (blue).

damaging agents (Fig. 5). Chk1 inactivates cdc25C through phosphorylation of serine 216 and consequent binding to 14-3-3 proteins. The cdc2–cyclin B1 complex is thereby prevented from becoming activated and initiating mitosis, and the cells arrest in G2. Although this arrest is initiated in cells without 14-3-3 σ or without p21, the arrest cannot be sustained. The 14-3-3 σ protein is normally

in a cytoplasmic complex with cdc2 during G2 arrest, and localization is dependent both on 14-3-3 σ (Fig. 4) and on continued nuclear transport by the *crm1*-dependent nuclear-export complex^{12–14}. In the absence of either 14-3-3 σ or nuclear transport, cdc2 and cyclin B1 escape to the nucleus and initiate a catastrophic mitotic process⁴. Thus, two different 14-3-3 proteins appear to

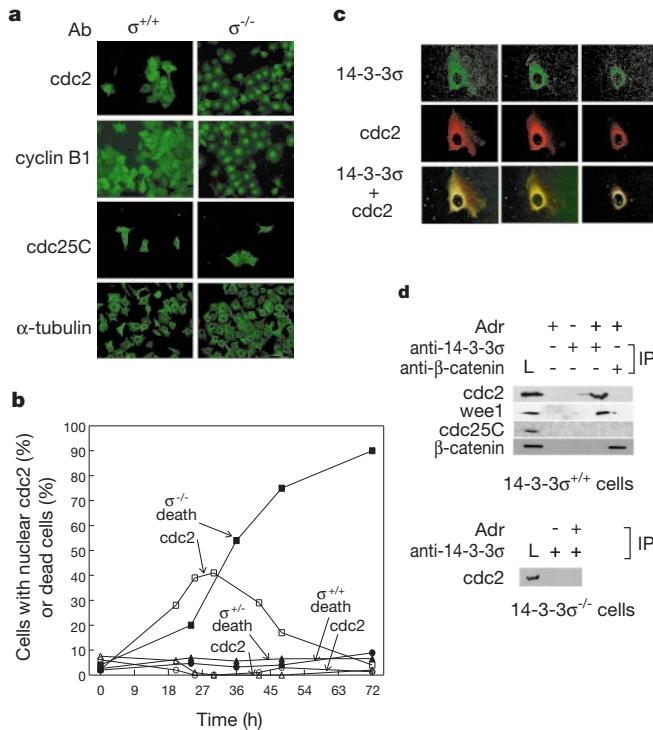


Figure 4 14-3-3 σ -deficient cells cannot sequester cdc2 and cyclin B1 in the cytoplasm following DNA damage. **a**, Cells of the indicated genotypes were treated with adriamycin for 48 h and stained with the indicated antibodies (Ab). **b**, Time course of nuclear migration and mitotic catastrophe of cdc2 in cells of the indicated genotypes treated with adriamycin. Cells at the indicated time points were scored for either mitotic death or mitotic catastrophe (indicated by chromatin condensation and micronucleation). **c**, Cdc2 and 14-3-3 σ colocalization. Cells were treated with adriamycin for 40 h and stained with antibodies to 14-3-3 σ (green) and cdc2 (red) before confocal microscopy. The panels show different planes through the same cell. Colocalization is indicated by the yellow colour in the bottom panels. **d**, Immunoprecipitation and western blotting. In the top panel, parental cells were treated with adriamycin (Adr) for 40 h in the lanes indicated. Lysates were prepared and immunoprecipitation was carried out using antibodies to 14-3-3 σ or β -catenin (a control) (see Methods). Immunoprecipitated proteins were separated by SDS-PAGE, transferred to membranes and probed with the indicated antibodies. The lanes marked 'L' contain aliquots of the crude adriamycin-treated cellular extracts. Top panel, results obtained using lysates of 14-3-3 $\sigma^{+/+}$ cells; bottom panel, results using lysates of 14-3-3 $\sigma^{-/-}$ cells.

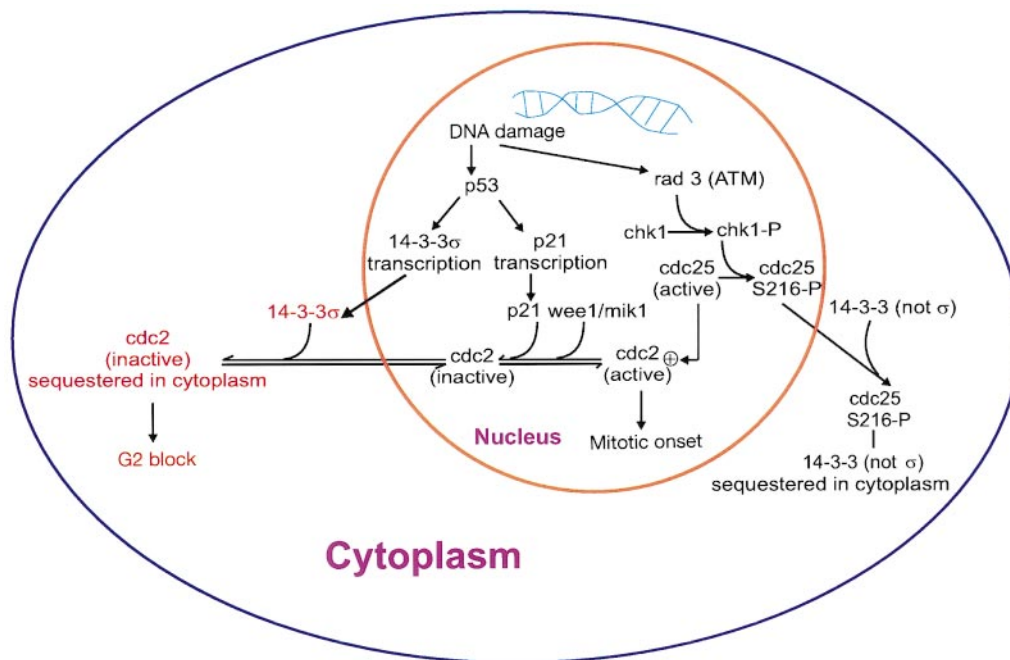


Figure 5 Model for G2/M checkpoint compartmentalization. The G2/M checkpoint is initiated by the phosphorylation of chk1 by rad3 family members (such as ATM in mammalian cells)²⁰. Chk1 then inactivates cdc25C through phosphorylation of serine 216, which leads to the binding of cdc25C to a 14-3-3 protein (not 14-3-3 σ) and to its exportation out of the nucleus¹⁵. The cdc2–cyclin B1 complex is thereby prevented from

becoming activated and initiating mitosis, and the cells arrest in G2. DNA damage also leads to stabilization of p53, which is required for maintenance of the G2 arrest through the transactivation of the *p21* and *14-3-3 σ* genes. 14-3-3 σ is required to sequester cdc2–cyclin B1 complexes in the cytoplasm and prevent mitotic catastrophes. p21 may prevent any cdc2–cyclin B1 that enters the nucleus from becoming activated.

ensure that mitosis does not occur in the presence of DNA damage, one by sequestering cdc25C¹⁵ and the other by sequestering cdc2–cyclin B1. It is becoming apparent that it is not only the enzymatic activities of cyclin-dependent kinase complexes but also their spatial compartmentalization that is critical for proper control of the cell cycle⁵. □

Methods

14-3-3 σ targeting construct

A BAC clone containing 14-3-3 σ was obtained as described¹. The BAC clone was digested with *Bam*HI, and two fragments, one 4.7 kilobases (kb) and the second 3.8 kb, were used to construct the 5' and 3' arms of the targeting vector, respectively. The 4.7-kb subclone contained the region immediately 5' of the initiating codon of the 14-3-3 σ coding region. The 3.8-kb subclone contained a region beginning 900 base pairs (bp) distal to the initiating codon. A synthetic *Eco*RV site plus the sequence 5'-CGTGGAGAGGGACTGGCAG-3', derived from a region of the 14-3-3 σ locus deleted by the targeting construct, were ligated to the 5' end of a geneticin-resistance gene. This product was placed into the pBluescript plasmid (Stratagene). To facilitate knockout of both alleles, *loxP* sites surrounding the geneticin-resistance gene were incorporated into the vector. Details of the constructs are available upon request. The construct was linearized by digestion with *Not*I and used for transfection. Southern blot analysis was carried out using standard techniques.

Cell culture and transfection

HCT116 cells were obtained from the American Type Culture Collection (ATCC). The HCT116 p21^{-/-} line has been described³. Cells were cultured in McCoy's medium supplemented with 10% fetal bovine serum (Gibco). We carried out transfections with Lipofectamine as directed by the manufacturer (Gibco). Clonal selection following transfection with the knockout construct was carried out in McCoy's medium with 10% fetal bovine serum and 0.4 mg ml⁻¹ geneticin (Gibco). Following transfection, cells were diluted in selection media and plated out in 96-well plates. After selection, we prepared genomic DNA from the drug-resistant clones using the QiaAmp column system (Qiagen). Identification of clones with successful targeting events was achieved using a PCR-based screen with the primers 5'-AGTGTCTGGGATCTCCAGC-3' and 5'-CTGCCAGTCCCCTCTCCAGC-3' and Taq Platinum (Gibco). PCR products were resolved by electrophoresis in 1% agarose gels. Knockout clones were confirmed by Southern blot and western blot analysis. Irradiation was performed using a 137Cs γ -irradiator at 1 Gy min⁻¹ for 12 min. Adriamycin was used at a concentration of 0.2 μ g ml⁻¹ (ref. 16). Leptomycin B was a gift from M. Yoshida. Time-lapse microscopy was done as described¹⁷.

Flow cytometry

Cells were trypsinized, washed with HBSS (Gibco) and resuspended in 40 μ l HBSS. The cells were then added to 360 μ l of a solution containing 1% NP-40 (Sigma), 4.7% formaldehyde (J. T. Baker), and 11 μ g ml⁻¹ Hoechst 33258 in PBS. The fixed and stained cells were stored at 4 °C and analysed within three days by flow cytometry.

Immunoprecipitation and western blot analysis

Immunoprecipitations were performed as described^{9,18}. 14-3-3 σ antibody was conjugated to beads using an Affi-Gel kit (Bio-Rad). Samples of protein from equivalent numbers of cells were fractionated on SDS-polyacrylamide gels (Novex). The proteins were then transferred to Immobilon P (Millipore) and incubated with either the p53-specific DO1 monoclonal antibody, the p21-specific EA10 monoclonal antibody (Calbiochem), a 14-3-3 σ -specific polyclonal antibody, anti-cdc2 polyclonal antibody (Santa Cruz), anti-wee1 antibody (Santa Cruz), or anti-cdc25C antibody (Santa Cruz). Polyclonal antibodies against 14-3-3 σ were generated in rabbits using a KLH-conjugate of the peptide SNEEGSEKGPVEV and were affinity purified using the same conjugate. The membranes were then washed with phosphate buffered saline (PBS) and incubated with the appropriate horseradish-peroxidase-coupled secondary antibody (Jackson Labs). Proteins were visualized with ECL (Pierce).

Immunohistochemistry

Cells were rinsed twice with PBS and then fixed with Histochoice (Amresco), permeabilized with 1% NP40 in PBS, and blocked in goat serum for 1 h. Antibodies specific for cdc2, cyclin B, cdk4, cyclin D, α -tubulin (all from Santa Cruz), 14-3-3 σ (generated as described above) and p21 (Calbiochem) were applied in GT (goat serum containing 0.05% Tween 20). After washing in PBST (PBS with 0.05% Tween 20), fluorochrome-labelled secondary antibody (Molecular Probes) was applied in GT for 1 h. Cells were washed for 5 min in PBST 3 times. In some cases, cells were counterstained with 4,6 diamidino-2-phenylindole (DAPI). Slides were mounted in DAPCO/glycerol and analysed with a Nikon Eclipse E800 microscope equipped with a CCD camera (Photometrics). Images were pseudocoloured using the software program IPLab (Signal Analytics Cooperation). TUNEL analysis was performed as described¹⁹. Cdc2 kinase assays were performed as described¹⁷.

Received 28 June; accepted 2 August 1999.

- Hermeking, H. *et al.* 14-3-3 σ is a p53-regulated inhibitor of G2/M progression. *Mol. Cell* **1**, 3–11 (1997).
- Busler, D. E. & Li, S. W. Rapid screening of transgenic type II and type XI collagen knock-out mice with three-primer PCR. *Biotechniques* **6**, 1002–1004 (1996).
- Waldman, T., Kinzler, K. W. & Vogelstein, B. p21 is necessary for the p53-mediated G1 arrest in human cancer cells. *Cancer Res.* **55**, 5187–5190 (1995).
- Heald, R., McLoughlin, M. & McKeon, F. Human wee1 maintains mitotic timing by protecting the nucleus from cytoplasmically activated cdc2 kinase. *Cell* **74**, 463–474 (1993).
- Pines, J. Cell cycle: checkpoint on the nuclear frontier. *Nature* **397**, 104–105 (1999).
- Nurse, P. Universal control mechanism regulating onset of M-phase. *Nature* **344**, 503–508 (1990).
- Hagting, A., Karlsson, C., Clute, P., Jackman, M. & Pines, J. MPF localization is controlled by nuclear export. *EMBO J.* **17**, 4127–4138 (1998).
- Jin, P., Hardy, S. & Morgan, D. O. Nuclear localization of cyclin B1 controls mitotic entry after DNA damage. *J. Cell Biol.* **141**, 875–885 (1998).
- Honda, R., Ohba, Y. & Yasuda, H. 14-3-3 zeta protein binds to the carboxyl half of mouse wee1 kinase. *Biochem. Biophys. Res. Commun.* **230**, 262–265 (1997).
- Peng, C. Y. *et al.* Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science* **277**, 1501–1505 (1997).
- Gardner, R., Putnam, C. W. & Weinert, T. RAD53, DUN1 and PDS1 define two parallel G2/M checkpoint pathways in budding yeast. *EMBO J.* **18**, 3173–3185 (1999).
- Toyoshima, F., Moriguchi, T., Wada, A., Fukuda, M. & Nishida, E. Nuclear export of cyclin B1 and its possible role in the DNA damage-induced G2 checkpoint. *EMBO J.* **17**, 2728–2735 (1998).
- Fukuda, M. *et al.* CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308–311 (1997).
- Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, J. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).
- Lopez-Girona, A., Furnari, B., Mondesert, O. & Russell, P. Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein. *Nature* **397**, 172–175 (1999).
- Gardner, R., Lengauer, C., Kinzler, K. W. & Vogelstein, B. Uncoupling of S phase and mitosis induced by anticancer agents in cells lacking p21. *Nature* **381**, 713–716 (1996).
- Bunz, F. *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497–1501 (1998).
- Reynisdottir, I., Polyak, K., Iavarone, A. & Massague, J. Kip/Cip and Ink4 Cdk inhibitors cooperate to induce cell cycle arrest in response to TGF- β . *Genes Dev.* **9**, 1831–1845 (1995).
- El-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
- Nurse, P. Checkpoint pathways come of age. *Cell* **91**, 865–867 (1997).

Acknowledgements

We thank the members of the Vogelstein/Kinzler Laboratory for helpful discussions. We thank M. Yoshida for the gift of leptomycin B and D. Tomasallo for helpful advice. Under an agreement between CalBiochem and Johns Hopkins University, K.W.K. and B.V. are entitled to a share of the sales royalty for the p21 antibody received by the University from CalBiochem. The terms of these arrangements are managed by the University in accordance with its conflict of interest policies. This work was supported by the Clayton Fund, the Medical Scientist Training Program and the NIH.

Correspondence and requests for materials should be addressed to B.V. (e-mail: vogelbe@welchlink.welch.jhu.edu).

Better safe than otherwise

Safety first, second and twelfth in the gleanings from recent new products.

Biocontainment rotor

Life Sciences International

www.lifesciences.com

A biocontainment rotor for Micromax microcentrifuges

A double-sealed biocontainment rotor is now available for LSI's Micromax range of microcentrifuges. The rotor can be easily removed from the centrifuge itself thus allowing opening under an environmental hood. The polypropylene rotor is fully autoclavable, or can be decontaminated using bleach and agents such as Count-Off without the risk of damage.

Reader Service No. 100

Pipette filter

Socorex

www.socorex.ch

Protection when pipetting hazardous liquids

The Acura 831, 0.5–5 ml, Calibra 832, 0.2–2 ml and 1–10 ml macropipettes from Socorex are fitted with a protection filter in the pipette nozzle. Made of cellulose, the filter prevents liquid or aerosol particles from entering the pipette barrel, eliminating risk of contamination. The filter is easy to remove or replace.

Reader Service No. 101

Hydrogen gas

Whatman International Ltd

www.whatman.com

Hydrogen gas on tap

This high flow Hydrogen Gas Generator is specifically designed to eliminate the need for using high pressure hydrogen gas cylinders in the laboratory. The Whatman Type 75-36 Generator is capable of continuously producing 550 ultra-pure hydrogen gas (at 99.99999+%) at a rate of 550 ml per minute



Gas meter: Whatman's steady supply of hydrogen.



Self contained rotors for Micromax centrifuges.

at regulated pressures from 0 to 4 barg, eliminating the need to change cylinders.

Reader Service No. 102

Biosafety cabinet

Jouan

www.jouaninc.com

This cabinet is alarmed

Jouan's MSC.12 Class II Type A/B3 Biosafety Cabinet features painted interior workspace walls to eliminate the glare associated with the more common stainless steel interior. The laminar airflow in the cabinet is measured and controlled by a microprocessor. An alarm sounds if the airflow is disturbed and alarms for low and high airflow, plugged filter, and power off are standard. The cabinet is produced in two 48-inch bench-top models. The basic model is supplied with one 115 V duplex outlet, the second has two, plus 2-gas/vacuum service valves.

Reader Service No. 103

AS-10 AgarMatic

New Brunswick Scientific

www.nbsc.com

Operator safety is the priority for this agar sterilizer

Safety features of the AS-10 AgarMatic include a lid that cannot be opened until the media have cooled to 5–10 °C and safety relief valves on the vessel and water jacket venting directly into the drain to avoid hot agar and water splashes. To protect the operator from injury when injecting additives to high temperature contents, the addition port has a thermal isolator.

Reader Service No. 104

Vacutainer CPT

Becton Dickinson

www.vmdp.bd/cpt

New information on this secure blood collection system for clinical trials

Vacutainer CPT is a one-step, closed system for blood collection, cell separation and transportation. The evacuated Vacutainer

CPT contains all the elements required for mononuclear cell separation in preparation for assays in multi-site clinical studies. The system was first introduced in 1994, but the manufacturers have published new data on the system in use. Call (in US) 1-800-631-0174 Option 3 for reports on mononuclear cell isolation for HLA typing; stability of plasma culture for HIV, RNA and p24 antigen in Vacutainer tubes; and lymphocyte proliferation assays for HIV-infected patients.

Reader Service No. 105

Zymit

International Products Corporation

Sample this enzyme-based cleaner free

This multipurpose enzyme-based cleaner called is designed to remove 'difficult' biological contaminants. Zymit is low-foam and mild (pH=7.5), made from enzymes and nonionic surfactants. The enzymes hydrolyse the proteins and starches, and the surfactants wash them away. Zymit is effective on food, grass, sweat, blood, fat, mucus, and feces. Contact the manufacturers for a free sample.

Reader Service No. 106

Disposable respirators

Sentinel Laboratories Ltd

Breathe easy in dusty environments

Protection from dust, mist, fumes and fibres is provided by the Respair range of respirators. These are disposable, avoiding the need for maintenance and increasing hygienic



Face valued: bring on the dust.

protection because each mask has only one wearer. The range is comprised of six types of masks including three 'valved' models, the P1, P2 and P3, with an exhalation valve to increase comfort. This valve also prevents the misting-up or 'fogging' of safety spectacles where applications require eye protection in addition to breathing protection. **Reader Service No. 107**

Bioprobe luminometer

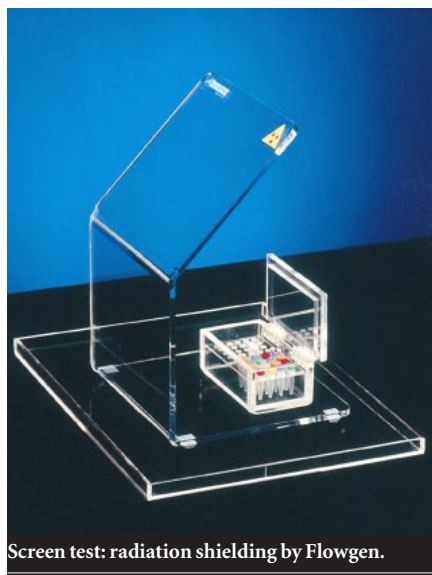
Pall Gelman www.pall.com/gelman
Portable ATP testing of microbial contamination

The BioProbe Luminometer detects the presence of microorganisms on filters, can perform direct measurements on work surfaces and detect microorganisms in liquids. It works by measuring ATP as the proxy for microbial contamination. Results appear immediately on a digital readout, can be printed out or stored for transfer to a PC. Results are recorded as counts per second/relative light units or as bacteria. **Reader Service No. 108**

Radiation protection

Flowgen www.flowgen.co.uk
A new range of beta and gamma radiation protection products

This new range of Beta and Gamma radiation



Screen test: radiation shielding by Flowgen.

protection includes radiation shields, workstations, spill trays, storage boxes and bins. Gamma radiation protection products are made from 12-mm thick optically transparent lead acrylic, equivalent to 0.5 mm lead. This range of products blocks gamma emissions from ^{125}I and any gamma emitters of less energy. It is not suitable for use with more energetic isotopes of iodine, but a 35-mm thick shield is available for use with isotopes of higher energy. Beta radiation protection products are made from 10-mm

thick optical acrylic, suitable for shielding beta emissions of ^{32}P , ^{35}S , ^3H and ^{14}C in biological research.

Reader Service No. 109

DuoFect

Quantum Biotechnologies Inc.
www.quantumbiotech.com

Nucleic acid delivery

Quantum specialize in Adenovirus Expression Systems, and now complement this viral gene delivery technology with DuoFect, a nucleic acid delivery system which makes use of transferrin receptor-mediated endocytosis to carry DNA into cells. The DNA is complexed with the cationic polymer polyethylenimine (PEI) to increase gene transfer efficiency.

Reader Service No. 110

Carboy Caddy

Labconco Corporation www.labconco.com

For transporting heavy water, or heavy anything

The Carboy Caddy is offered as a safe way of moving large carboys or cartons. The shelf is made of phenolic board with a Melamine high-pressure laminate surface which resists staining and corrosion. The shelf measures 24 43.3 × 61.3 cm. Casters (50 mm diameter) provide a low profile, elevating the shelf only 60 mm above the floor. The durable handle is made of 18-gauge corrosion resistant epoxy coated steel and measures wide (98 cm high × 48 cm wide). The Carboy Caddy is designed to negotiate narrow corridors and doorways and is delivered as a fully assembled 'runner'.

Reader Service No. 111

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.



You can handle it: a carboy takes to the road.

ADVERTISEMENTS

Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check Us Out

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com

New Probe Sizes
0.45cm!

READER ENQUIRY NO. 54

Diagnostic Reagents

Pefachrome®-Series: chromogenic and fluorogenic substrates

Pefabloc®-Series: serine proteinase inhibitors

Purified snake venom proteins (e.g. Protac®, Ecarin, Batroxobin)

PENTAPHARM LTD
CH 4002 Basel Fax: +41 61 319 96 19
<http://www.pentapharm.ch>

READER ENQUIRY NO. 16

Lysyl Endopeptidase
(Achromobacter Protease 1)

Specifically cleaves all Lys-X bonds including Lys-Pro.
Highly Stable, Water Soluble

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 23

CUSTOM PEPTIDES

• Up to 30 aa
• HPLC & MS data

Purity	>80%	>95%	>99%
Amount	15mg	10mg	

info@genosphere-biotech.com
Tel: +33 1 42 71 70 21 • Fax: +33 1 42 71 70 23

GENOSPHERE BIOTECHNOLOGIES

READER ENQUIRY NO. 24
© 1999 Macmillan Magazines Ltd